Annu. Rev. Physiol. 2004. 66:735–69
doi: 10.1146/annurev.physiol.66.082602.092845

First published online as a Review in Advance on October 20, 2003

INTERPRETING THE BOLD SIGNAL

Nikos K. Logothetis[1] and Brian A. Wandell[2]
[1]*Max-Planck Institut für Biologische Kybernetik, Tübingen, Germany*
[2]*Department of Psychology, Stanford University, Stanford, California;*
email: wandell@stanford.edu

Key Words visual cortex, blood-oxygen-level-dependent, fMRI, local field potential, multiunit activity

■ **Abstract** The development of functional magnetic resonance imaging (fMRI) has brought together a broad community of scientists interested in measuring the neural basis of the human mind. Because fMRI signals are an indirect measure of neural activity, interpreting these signals to make deductions about the nervous system requires some understanding of the signaling mechanisms. We describe our current understanding of the causal relationships between neural activity and the blood-oxygen-level-dependent (BOLD) signal, and we review how these analyses have challenged some basic assumptions that have guided neuroscience. We conclude with a discussion of how to use the BOLD signal to make inferences about the neural signal.

INTRODUCTION

Functional magnetic resonance imaging (fMRI) has greatly increased our ability to study localized brain activity in humans. Because humans are a cooperative and intelligent subject population, fMRI can be used to study problems that are very difficult to approach in animals. For example, what are the neural mechanisms that allow us to see objects, plan for the future, and recall ideas? How are feelings and thoughts represented in the brain?

The broad interest in these questions, coupled with the widespread availability of magnetic resonance (MR) scanners, brings together many investigators with differing perspectives on how to analyze brain computations. Also, each neuroscience tool, including the microscope, the microelectrode, and behavior, has limitations; each tool informs us about some, but not all, components of neural computation. Because of the diversity of investigators and the limited understanding of this new tool, fMRI data have been interpreted in a variety of ways.

As with other tools, fMRI will be integrated effectively into neuroscience when we can identify which aspect of the neural signals fMRI measures and how these measurements relate to those made with other instruments. Here, we describe our current understanding of the causal relationships between neural activity and the

blood-oxygen-level-dependent (BOLD) fMRI signals. We further analyze how to interpret the BOLD signal and indirect measure of brain activity in terms of the brain's synaptic activity and action potentials.

The process of understanding the relationship between the BOLD signal and neural events has challenged some basic assumptions that have guided neuroscience. For the past 40 years, measurements of cortical processing have been dominated by the analysis of action potentials. We now know that the BOLD signal does not correlate perfectly with action potentials, but rather measures a mix of continuous membrane potentials and action potentials. This complicates any comparison of it with single-unit physiology, but it also offers an opportunity to consider whether our understanding of neural information processing might extend beyond action potentials to include a range of signals that are an important part of neural computation.

With recent advances in understanding the physiological mechanisms that convert neural activity to BOLD signals, we have some answers about a question that is at the heart of the vast majority of experimental fMRI papers: What does the BOLD signal tell us about the neural signal? The reader should not be disappointed to learn that there are only partial answers as to how to design experiments that provide unambiguous interpretations from BOLD signals to neural signals. Methods of relating different neuroscience measurements have evolved over many decades in the areas of anatomy, microelectrodes, genomics, and behavior. Rather, we hope the reader will be encouraged to learn that good progress has been made and that it is possible to seriously address this important issue. The main portion of the review describes our view of the signaling pathway.

We conclude with a speculative discussion of what the BOLD signal can teach us about the neural signal. We note that some commonly used BOLD measures, for example the absolute size of the fMRI contrast signal, cannot be relied upon to measure the amplitude of the neural responses at two different cortical locations. On the other hand, measurements of stimulus selectivity at a single cortical location have a better logical foundation. Measurements of spatial maps are on a solid footing as well, although the spatial precision of map measurements may differ across cortex. We describe a preliminary mathematical framework that may be useful when extrapolating from BOLD to neural signals.

THE BOLD SIGNAL

The MRI Signal

Magnetic resonance measures how radio frequency electromagnetic waves act upon dipoles in a magnetic field. In the specific case of brain measurements, the MR signals arise mainly from the hydrogen nuclei in water, which are the only dipoles present in significant density to support measurements at high spatial resolution. The MR signal measures how these dipoles transition between different energy states. The now familiar MR images represent measurements of how these

transitions differ depending upon the properties of the nearby tissue or the physiological state of the brain.

The hydrogen nuclei achieve a relatively organized, low-energy state when the subject enters the static (B0) field of the MR scanner. The strength of this magnetic field is specified in units of Tesla, and common field strengths range from 1.5 to 4.7 T. In the presence of a strong magnetic field, there is a tendency for the magnetic moment of these nuclei to align parallel or antiparallel to the main field—in much the same way as the dipole moment of a small bar magnet aligns itself with the local magnetic field.

The magnetic resonance measurement begins when the experimenter introduces a radio frequency (rf) pulse into the tissue. This pulse excites nuclei away from their resting state into a higher energy state. For any particular dipole, such as a hydrogen nucleus, this excitation is only effective at a resonance frequency known as the Larmor frequency. At this frequency, which is proportional to the magnetic field strength and the gyromagnetic constant of the hydrogen nucleus, the radio pulse contributes precisely the amount of energy needed to cause the nuclei to transition to a higher energy level and realign in the magnetic field.

The rf excitation pulses and the magnetic field gradients superimposed on the B0 field can be applied according to a variety of different timing and amplitude parameters. Adjustments to these parameters permit the investigator to obtain images that distinguish various properties of the brain, including structure (anatomical imaging), flow (perfusion imaging), or neural activity (functional imaging).

Information about the nearby tissue is derived from the rate at which the hydrogen nuclei return to the low-energy state following the excitation. Imagine the average initial orientation of a population of dipoles as a small vector pointed in three-dimensional space; this vector is aligned with the B0 field. The rf excitation pulse changes this population average. The relaxation back to the original state can be described as changes in two dimensions, longitudinal re-growth and transverse relaxation. Two exponential processes with time constants (T1 and T2) describe the relaxation back to the low-energy state. The T1 constant measures the relaxation in the direction of the B0 magnetic field (longitudinal re-growth). The T2 constant measures the transverse relaxation of the dipole in the *x-y* plane that is perpendicular to the B0 field. These changes in the local magnetic field are measured by special equipment (coils) that is placed within the scanner.

The transverse relaxation, also called spin dephasing or spin-spin interactions, is of special significance for fMRI. Any energy transition of a nucleus changes the local field at nearby nuclei; these exchanges are called spin-spin interactions. In an ideal homogeneous magnetic field, the transverse relaxation follows an exponential signal decay (free-induction decay, FID); the time constant of the decay is called T2. However, in physiological tissue the transverse relaxation is more rapid because of local field inhomogeneities including those caused by the tissue itself. When inhomogeneities are present, the decay constant is called T2*. Field variations randomly alter the frequency of the proton's precession, disturbing the

phase coherence and speeding the transverse relaxation. In the brain, the size of these inhomogeneities depends upon the physiological state and in particular the composition of the local blood supply. This physiological state depends, in turn, on the neural activity. For this reason, measurement of the T2* parameter is an indirect measurement of neural activity.

MR signals are significant for neuroscience applications because it is possible to measure T2* at fairly high spatial resolution across the entire brain. These measurements are obtained by superimposing small gradients upon the main B0 magnetic field (for descriptions on image formation, see 1, 2). Much of the present review describes the chain connecting the neural activity and the resulting spatially varying field inhomogeneities that are measured by T2*.

The BOLD Contrast Mechanism

The mechanisms connecting neural activity to the measured T2* value are generally called the BOLD contrast mechanism; this measurement is currently the mainstay of brain fMRI studies.

As its name suggests, the BOLD contrast mechanism alters the T2* parameter mainly through neural activity–dependent changes in the relative concentration of oxygenated and deoxygenated blood. Deoxyhemoglobin (dHb) is paramagnetic (3) and influences the MR signal (4) unlike oxygenated Hb. In the presence of dHb, the T2 value decreases quadratically with field strength, as expected from the dynamic averaging owing to diffusion in the presence of field gradients (5, 6). The effects of dHb on T2* are even stronger, as first noticed by Ogawa et al. in their seminal studies on the rat brain in high fields (7–9). Specifically, Ogawa & Lee observed that blood vessel contrast varied with changes in blood oxygen demand or flow (9). They attributed the contrast increase to a magnetic susceptibility effect associated with the paramagnetic deoxyhemoglobin in red cells (7). They recognized the significance of their finding and concluded "BOLD contrast adds an additional feature to magnetic resonance imaging and complements other techniques that are attempting to provide positron emission tomography-like measurements related to regional neural activity" (8). Subsequently, the effect was demonstrated in the cat brain during the course of anoxia (10).

Given these observations, the reader may now be wondering about the origin of the signal increases typically reported in imaging experiments. Upon neural activation, any increases in dHb would be expected to enhance the field inhomogeneities and reduce rather than increase the BOLD contrast. The observed enhancement is due to an increase in cerebral blood flow (CBF) that overcompensates for the decrease in oxygen, delivering an oversupply of oxygenated blood (11, 12). At present, the reason for the mismatch between supply and consumption of blood oxygen is unclear. In a perfectly regulated system, the blood oxygen level would follow demand as neural activity waxed and waned. In fact, glucose supply does appear to match the consumption. Why should the glucose supply, but not the oxygen supply, match demand?

One possibility is that the vasculature delivers a fixed ratio of oxygen and glucose that is appropriate for an aerobic process. If both aerobic and anaerobic processes demand glucose, then the result would be an oxygen surplus (e.g., see 13). According to this hypothesis, the intensity of the BOLD signal depends on the relative proportion of local aerobic and anaerobic glucose metabolism, and understanding the distribution of anaerobic metabolism becomes an important consideration in interpreting the BOLD signal.

An alternative hypothesis is that oxygen extraction matches the metabolic needs and the excess oxygen present in the blood supply is due to an inefficient delivery process. Specifically, it has been proposed that this oversupply compensates for the inefficient, passive oxygen diffusion that occurs at high flow rates (14, 15). On this hypothesis, oxygenation supply is tightly coupled to neural activity, and the anaerobic processes in the astrocytes represent only a negligible amount of energy that is supplied by existing reserves (16, 17). There are some contradictory observations, however, so that the interpretation of these circulatory and metabolic changes remains uncertain (17, 18).

Physiological explanations of the BOLD signal can be tested by comparing theory with the amplitude and time course of the measured BOLD. The time course of the human BOLD response to a brief stimulus, the temporal impulse response function, is often called the hemodynamic response function (HRF). There is heterogeneity in the HRF across cortex of an individual observer (or animal), between observers, and possibly across different sensory, motor, and cognitive tasks (19–22). Figure 1*a* shows examples of the time course of the BOLD signals in visual and motor cortex after subjects viewed a 2-s stimulus or performed a motor act lasting 2 s (19, 23). The neural response to the sensory and motor acts ends fairly quickly, within a few hundred milliseconds past $T = 0$; the BOLD response begins roughly 2 s later, e.g., (24), rises to a plateau 6–9 s after stimulus onset, and then returns to baseline. In some instances the BOLD response has a post-stimulus undershoot, as can be seen in one of the four curves in Figure 1a (25–27).

There has been considerable focus on whether linear systems methods accurately model the BOLD signal. Specifically, a number of investigators have tried using the responses to brief stimuli (2 s) to predict the responses to visual or motor acts of longer duration (>8 s). Examples of measurements of responses to 8 s of visual stimulation and an 8-s motor act are shown in Figure 1*b*. The simple prediction of a linear time invariant (LTI) system, in which the response to the long-duration event is predicted from that to the short-duration event, fails significantly (19, 28, 29). It is possible, however, to make reasonably accurate predictions for responses to long-stimulus presentations on the basis of responses to stimuli of 6 s and greater (19, 28, 30) and to predict responses to brief stimuli from the responses to even briefer stimuli. Several authors have proposed modifications to the linear model to account for the inability to predict very long-duration responses from very short-duration responses (19, 28).

Despite the imprecision of the linear model predictions, a number of theorists have proposed formulae to describe the impulse response function. The formulae

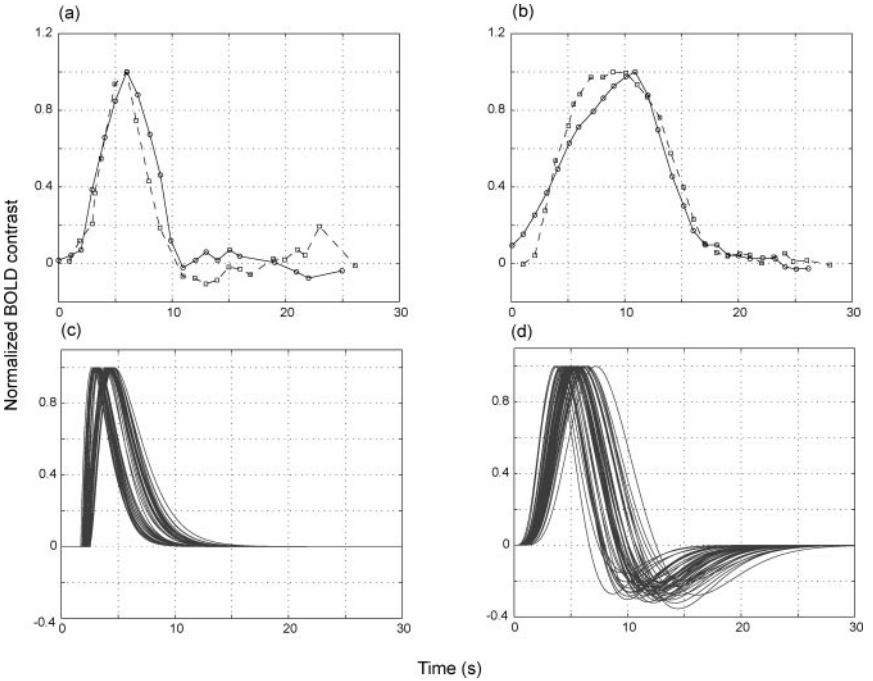

Figure 1 Time course of the BOLD response. (*a*,*b*) Data are replotted from experiments in motor cortex (*open circles*) (19) and visual cortex (*open squares*) (161). The two panels show measurements in response to a visual stimulus or movement of 2 s (*a*) or 8 s duration (*b*). (*c*,*d*) Theoretical temporal impulse response functions (also called hemodynamic response functions) used in the statistical analyses of the BOLD response. These curves are intended to represent the BOLD response to brief stimuli of unit amplitude. The distributions of curves shown in these figures are derived from formulae in (162) (*c*) and (32) (*d*). The distributions of curves were created by randomly perturbing the parameters in the respective formulae by 10%.

for the HRF usually include parameters that permit the investigator to approximate a range of HRFs such as those in Figure 1*a*,*b*. Data analysis packages all include some assumption about the HRF, and two examples are shown in Figure 1*c*,*d*. The collection of curves in Figures 1*c*,*d* show the consequence of varying the parameters over a modest (10%) range (30–32).

The Significance of Human fMRI

Ogawa's groundbreaking work generated great interest in human BOLD fMRI. In 1992, three groups simultaneously and independently took the method a step further by measuring the intrinsic BOLD signal in humans (24, 33, 34). These reports initiated the flood of fMRI publications that have appeared in scientific journals ever since.

The success of fMRI stems partly from the availability of scanners, the noninvasive nature of the measurement, and its high spatiotemporal resolution compared with other methods available for human research. Prior to the BOLD fMRI, the only methods available to study the alert human brain were electro-encephalograms (EEG; scalp potentials), magneto-encephalograms (MEG), and positron emission tomography (PET). These methods suffer from either poor resolution or lack of spatial localization. The two-point spatial resolution of fMRI, on the other hand, at least in visual cortex, is on the order of two millimeters (35). Increased averaging can, in principle, yield comparable spatial resolution in PET, but this method requires the introduction of radioactive substances into the subject so that extensive averaging or even many repeated studies of a single brain are not practical. Using fMRI, one can study the same human brain for many hundreds of hours without significant health risk. This ability both improves the signal-to-noise ratio and permits the study of individuals whose brains are of particular interest, such as neurological case studies.

The Neural Signal

To form a more complete picture of the neural mechanisms underlying behavior, results from human fMRI experiments are often compared with those obtained in electrophysiology made in homologous regions of monkey cortex. The vast majority of such experiments in conscious animals report extracellular electrophysiological recordings. It is therefore worth considering the question: What do extracellular electrodes measure?

The Extracellular Field Potential

The extracellular medium surrounding neurons is a volume conductor with a resistivity (specific impedance) that ranges from 200–400 Ωcm depending on the neural site (36–39). This extracellular resistivity is higher than that of a saline bath, ~65 Ωcm, 1 Hz to –10 kHz (N.K. Logothetis, unpublished measurements). The increased resistivity is due to limitations on the flow of ions in the extracellular medium compared with the saline. For the frequency range of electrophysiological signals (0 to about 2 kHz), inductive, magnetic, and propagative effects can be ignored, and the electric field can be considered homogeneous and purely resistive, obeying Ohm's law. In such a static description, the flow of positive ions (e.g., Na^+) into the active sites of a neuron appears as a current sink (inward current); the injected current flows down the core of the dendrites or axon. Because currents flow in closed loops, a current outflow from distant inactive membrane sites appears as a source (outward current) from which current flows back to the injection site through the extracellular medium. The finite resistance of the latter creates the extracellular field potentials (EFPs) that are measured by electrodes.

The weighted spatio-temporal sum of the sinks and sources from multiple cells is often called the mean extracellular field potential (mEFP; the black trace in

Figure 2*a*). The weighting depends on several factors. Most importantly, the spatial alignment of elongated neural processes establishes an anisotropic conductivity across space and imparts a preferred directionality to the current flow. The directionality complicates the interpretation of the mEFP measurement: Cells whose current flow travels in opposing directions will cancel each other out. There are no practical methods for recovering information about the spatial conductance pattern because such information is rarely available, and there is no widely agreed upon model for interpreting the mEFP with respect to the factors that influence the cellular currents.

Despite its limitations the mEFP provides an important tool for system analysis. Depending on the recording site, on the choice of electrode size and type, on signal processing, and on the sophistication of mathematical analyses of its spatiotemporal distribution, the mEFP can be still used to explore the functional properties of neural tissues at a large range of spatial scales. It can be employed to study properties ranging from single neurons to neuronal ensembles on a scale of hundreds of microns. In neural structures with known anatomical organization the study of ensembles often provides insights in the cellular components involved in different computations, and the comparison of single- and multiple-unit activity is useful for defining functional modules. We suspect the same will prove true of the fMRI signal, especially as we learn more about how to constrain its interpretation by combining imaging with other invasive or non-invasive modalities.

Spiking Activity

If a microelectrode with a small tip is placed close to the soma or axon of a spiking neuron, spikes can be detected in the mEFP signal (Figure 2*b*). It is also possible to measure spikes simultaneously from a collection of cells in a small neighborhood using arrays of electrodes (40). Recent studies in rats, for instance, show that a group of four closely spaced electrodes (tetrodes) placed within 50–100 μm of pyramidal neurons in hippocampus provide accurate information on a number of spike parameters, including latency and amplitude; these spikes are measured simultaneously by intracellular recordings (41, 42).

Ever since the early development of microelectrodes, action potentials of well-isolated neurons have been used to understand the neural basis of behavior (43). The dominance of the single-unit action potential was strengthened by its suitability for studying the first steps of afferent information processing in sensory physiology (44–47).

Computation

There are compelling reasons for focusing on the role of the action potential; for example, the action potential is the only way neurons communicate over a distance. Hence, ganglion cell action potentials contain all of the information communicated from the retina to the rest of the nervous system. In general, information between major structures—such as the LGN to V1—is contained within the action potentials

of the output neurons of these structures. Understanding this representation is essential to understanding neural computations.

Put this way, it is also easy to see the drawback of studying only action potentials: We learn little about the sub-threshold integrative processes that are essential for producing spikes within these structures. Within the retina, for example, ganglion cells are the only spiking neurons. If we restrict our study to ganglion cell spikes, we would miss the origins of trichromacy in the continuous photo-responses of the cone photoreceptors; we would miss how the rod and cone circuitry combine to represent signals spanning eight orders of magnitude; the formation of color opponency and the organization of the rod pathways would stay hidden from view. An analysis of the retina restricted to spikes is superficial, much like a view of the cortex that ends by declaring one region a face area and another a color area. Measuring only the spikes on large axons projecting between areas misses an essential question: What is the complete set of neural computations, including spikes and sub-threshold synaptic potentials, that create the output responses?

Sampling Bias

Single-unit recordings, in particular using large electrodes, do not sample cell sizes and types with equal probability (48, 49). Electrodes are more likely to sample large cells, and this bias, in turn, is partially responsible for the cell-type bias. The reason for the bias can be understood from basic signaling considerations. An action potential from a large neuron generates greater membrane current and a larger extracellular spike than an action potential from a small cell. Consequently, the extracellular field from a large cell remains above recording noise levels over a greater distance (50). For distances greater than about 140 μm, spikes become indistinguishable from background noise (42). Consequently, microelectrodes are likely to sample preferentially the somas and axons of large cells, a prediction supported by experimental work (49, 51). Further, in awake-behaving preparations, movement artifacts are likely to shift the microelectrode out of the range of a small cell's extracellular potential, making it difficult to hold the signals from such cells for the duration of an experiment. For all these reasons, we suggest that extracellular electrodes probably over-sample large cells (e.g., pyramidal cells in cerebral cortex and Purkinje neurons in cerebellar cortex).

An interesting study by Henze and colleagues provides more evidence of sampling bias. These investigators made simultaneous in vivo, intracellular, and extracellular measurements in the hippocampus of the anesthetized rat. They placed a tetrode array within the hippocampus and repeatedly advanced an intracellular electrode searching for a signal that could also be measured by the tetrode array. The 22 neurons that could be measured with the two types of electrodes were morphologically identified by a biocytin injection (42). Of these successful measurements, 21 were pyramidal cells and 1 was an inhibitory interneuron (42). This sampling distribution differs from any reasonable accounting of the distribution of cell types in hippocampus, which like the rest of cortex includes certain proportions of large- and medium-size pyramidal cells, different sizes of stellates, and

many inhibitory neurons as small as a few microns. The measured sample from the tetrode array suggests a strong bias either in the intracellular sample or in the ability to measure across cell types with the extracellular tetrode array.

Separating Sub-Threshold Potentials and Spiking Activity

If a microelectrode with a relatively large tip (low impedance) is placed in the extracellular space, somewhat distant from large spiking neurons, the mEFP is dominated by the analog synaptic voltages (dendritic events) and summed action potentials from hundreds of nearby neurons (52, 53). This mEFP voltage can be subdivided into two parts that measure either the dendritic events or the summed spikes (for review see 54, 55). Specifically, applying a high-pass filter (cut-off 300–400 Hz) measures the multiple-unit spiking activity (MUA; Figure 2*c*, *upper panel*); applying a low-pass filter (cutoff $<$ 200 Hz) produces a waveform that reflects the local synaptic voltages (LFP; Figure 2*c*, *lower panel*).

Multiunit Activity

Combined physiology-histology experiments demonstrate that the magnitude of MUA varies considerably across brain sites (e.g., neocortex versus hippocampus); within a region, however, the MUA magnitude is relatively constant. The local cell size is an important factor in determining the MUA magnitude (56, 57). Sites of large-amplitude and fast MUA signals generally have homogeneous populations of large cells (58). These experiments further show that amplitude of MUA signals covaries with the magnitude of extracellular spike potentials.

Local Field Potentials

The low-frequency range of the mEFP signal, the LFPs, represents slow electrical signals and sub-threshold activity. Until recently these signals were thought to represent synaptic events exclusively. Evidence for this came from combined electroencephalographic (EEG) and intracortical recordings showing that the slow-wave activity in the EEG is largely independent of neuronal spiking (59–62). Unlike the MUA, the LFP magnitude is not correlated with cell size. Instead, the LFP amplitude reflects the extent and geometry of dendrites in each recording site. Cells in the so-called open field geometrical arrangement, in which dendrites face in one direction and somata in another, produce strong dendrite-to-soma dipoles when they are activated by synchronous synaptic input. Other cortical neurons are oriented horizontally and contribute less efficiently or not at all to the sum of potentials. The pyramidal cells with their apical dendrites running parallel to each other and perpendicular to the pial surface form an ideal open field arrangement and contribute maximally to both the macroscopically measured EEG and the local field potentials.

Thus LFPs reflect primarily a weighted average of synchronized dendro-somatic components of the synaptic signals of a neural population, most likely from within

0.5–3 mm of the electrode tip (63, 64). Yet, it has been further shown that LFPs may also measure other types of slow activity unrelated to synaptic events, including voltage-dependent membrane oscillations (e.g., 65) and spike afterpotentials. To be more specific, the soma-dendritic spikes in the neurons of the central nervous system are generally followed by afterpotentials, a brief delayed depolarization, the afterdepolarization, and a longer lasting afterhyperpolarization, which are thought to play an important role in the control of excitation-to-frequency transduction (e.g., 66–68). Afterpotentials, which are generated by calcium-activated potassium currents (e.g., 67, 69–72), have a duration on the order of 10 s of milliseconds and most likely contribute to the generation of the LFP signals, as first suggested by Buzsaki and his colleagues (73, 74).

In summary, MUA measures regional neuronal spiking, whereas the LFP measures slow waveforms, including synaptic potentials, afterpotentials of somatodendritic spikes, and voltage-gated membrane oscillations. Hence, the LFPs are information-rich signals that may influence local neural excitations (75) and may reflect aspects of the input signal and the local intracortical processing mediated by the sub-threshold signals of interneurons. Unlike the MUA, the LFP does not reflect the action potentials carried by the principal (output) neurons.

BOLD IMAGES OF THE NEURAL SIGNAL

In the previous section, we reviewed a variety of ways to analyze the mEFP and described how each method could be an important measurement of the signals carried by the neuronal ensemble. In this section, we review the properties of the BOLD signal with the goal of finding its place within the array of measurements of the neural signal.

Sensory experiments frequently measure the relationship between stimulus energy and the BOLD response. In general, this relationship is nonlinear—the BOLD response increases according to a compressive, nonlinear, saturating function of stimulus energy (76). There are two stages of the path from stimulus energy to BOLD response that may introduce this nonlinearity: The neural signal may depend nonlinearly on the stimulus energy, and the BOLD response may depend nonlinearly on the neural signal. A number of studies using various techniques have presented data showing that in certain regimes there is a linear relationship between neural activity and the subsequent BOLD response (77–80). This hypothesis is important because if there is a linear relationship, there is a realistic possibility of inferring neural signals from the BOLD response.

Logothetis and colleagues recently examined the relationship between BOLD and neural activity by simultaneously acquiring electrophysiological and fMRI data from monkeys in a 4.7 T vertical scanner (27, 29, 81). The BOLD response reflected a local increase in neural activity as assessed by the mEFP signal. For the majority of recording sites, the amplitude of the BOLD response was a linearly increasing, but not a time invariant, function of LFPs, MUA, and the firing rate

of small neural populations. A proportional increase in neural and hemodynamic signals was observed in experiments varying (within a limited range) the stimulus luminance contrast (81). This linear relationship over the measured range did not extend to zero, suggesting that the entire relationship is nonlinear. The nature of the nonlinearity in the low-contrast ranges, specifically, whether it is the tail-end or saturation point of a compressive function as indicated to us by A. Movshon (personal communication), remains to be investigated. The correlation between firing rate and BOLD was also confirmed with MRI/MRS studies in rats, whereby a linear relationship between the cerebral metabolic rate, glutamatergic neurotransmitter flux, and neural activity was demonstrated (82, 83).

The relative contribution of MUA and LFP signals to the BOLD response was examined by applying time-dependent frequency analysis to the raw data (81). In all these experiments, increases in the LFP range were both of greater spectral power and higher predictive reliability. After stimulus presentation, a transient increase in power was typically observed across all frequencies, followed by a lower power level that was maintained for the entire duration of stimulus presentation. A prominent characteristic in all of the spectrograms was a marked stimulus-induced increase in the magnitude of the LFP; this increase was always larger than that observed for MUA. A decrease in neural firing rates was also observed immediately after the termination of the stimulus.

Comparing LFPs and MUA as Predictors of the BOLD Response

Although the BOLD response is not linearly dependent on neural signals under all conditions, linear time-invariance (LTI) is a reasonable first approximation over restricted ranges (54, 55, 84, 85). In an LTI system, responses to arbitrary stimuli can be predicted from knowledge of the input signal and the system's impulse response function.

To evaluate which of the neural signals is the causal agent in creating the BOLD response, Logothetis et al. examined how well LTI methods could be used to predict BOLD fMRI signal from LFPs, MUAs, and spikes (81). The neural and BOLD measurements were obtained from a variety of cortical sites. Measurements included simultaneous recording of visually driven signals as well as spontaneous brain activity. The latter signals, when additionally prewhitened, offer an unbiased estimate of the hemodynamic impulse response. Impulse response functions (see Figure 2) were estimated from the responses to briefly presented stimuli, and the estimated impulse response function was convolved with several types of neural signals to evaluate how well each type of signal predicted the BOLD response.

Figure 3 shows the neural responses as well as the measured and estimated BOLD responses for four stimulus durations at one cortical site. Estimates of the time course of the neural response were relatively accurate over a range of stimulus durations (up to 12 s). As the stimulus duration increases, so do the errors. Deviations from linearity of this sort have also been described in the

analysis of linear systems properties in the human literature (see above; e.g., 19, 30).

In general, LFPs and MUA vary in a similar manner. Hence, at those sites where the LFPs predicted the BOLD response, the MUA did too. Across cortical sites there was a tendency for the LFP-based estimate to perform slightly better than the MUA-based estimate: The LFP signal predicted 7.6% more of the variance than the MUA. The difference, although small, was statistically significant. The larger variability of MUA was mostly attributable to the stronger adaptation effects observed in this frequency range of the mEFP.

Spikes Can Be Dissociated from LFP Signals

There is a strong correlation between local field potentials and BOLD response. However, we still cannot be certain that the slightly greater predictive power of the LFP over the MUA implies that spikes are not a key variable. Because of the correlation between the LFP and MUA signals, either continuous potentials or spikes could be the key determinant of the BOLD response. Circumstances do exist, however, in which there is a dissociation between LFPs and spikes. Such a dissociation was observed in visually driven neuronal responses in striate cortex (81).

Figure 4 illustrates simultaneous measurements of LFP, MUA, and BOLD at a single cortical site in response to visual stimuli. The upper and lower panels show the responses to a 24- and 12-s stimuli, respectively. At this site, the MUA was of brief duration in response to both stimuli, returning to baseline approximately 2.5 s after stimulus onset. In contrast, the LFP duration was similar to that of the stimulus, remaining elevated for the entire stimulus duration. The BOLD response also varied with stimulus duration. Hence, at this site, where the MUA and LFP differ, only the LFP predicted the BOLD response. This dissociation could be observed in about 25% of the responses. There was no single observation period or recording site for which the opposite result was observed; that is, there was no measurement in which the MUA was correlated with the BOLD but the LFP was not.

An exquisite example of LFP-MUA dissociation was reported for cerebellar cortex by Lauritzen and collaborators (78, 86). In cerebellar cortex, climbing fiber stimulation causes monosynaptic excitation of the Purkinje cells, whereas simultaneous parallel fiber stimulation causes a disynaptic inhibition of the Purkinje cell spikes (via the basket cells). Lauritzen and colleagues measured LFPs, single-unit activity, and changes in cerebral flow by means of laser Doppler flowmetry (78). They demonstrated that despite the inhibition of Purkinje cell spikes, parallel fiber stimulation increases overall synaptic activity. Both LFPs and cerebral blood flow increased, following the synaptic activity, even though spiking activity ceased.

Finally, exploiting the differential effect of neuromodulators on different cellular sites may experimentally induce LFP-MUA dissociation. In preliminary results in macaques from the Logothetis laboratory, neuromodulators were injected during

simultaneous electrophysiological and BOLD neuroimaging experiments. Using a triple pipette (electrode, saline, and a neuromodulator), 20 microliters of 0.01 mol 5HT (5-hydroxytryptoamine hydrochloride) were injected over a period of 10 min. Several minutes after the injection a profound suppression of the MUA was observed. The LFP signal showed a slight increase and returned to baseline within a few minutes. During this time, in which the MUA was silenced, no significant change was discernible in the BOLD response. Spectrograms obtained before and after the 5HT injection during visual stimulation confirmed that the stimulus-induced spikes were entirely eliminated, although LFP activity was moderately increased. These measurements show that it is possible to dissociate pharmacologically spiking activity and hemodynamic responses. On the basis of the several dissociations described in this section, we conclude that the LFP signal is the key variable for the BOLD response.

In summary, the BOLD response primarily reflects the input and local processing of neuronal information rather than the output signals, which are transmitted to other regions of the brain by the principal neurons. The long-range projection signals from these principal neurons are the measurements that are mainly accessible in single-cell recordings in the behaving animal.

NEURAL SIGNALS, METABOLISM, AND BLOOD FLOW

In the previous section, we reviewed the correlation between the neural signal and the BOLD response. Here we examine the cellular and molecular neurovascular coupling mechanisms that mediate this correlation. These mechanisms are the scaffolding for the neurometabolic coupling between energy demand and supply.

The brain, as does the heart, makes ongoing demands for energy. The mean level of energy required by the brain is very high, comparable to the energy required by the heart. On top of this mean level, the regional demand fluctuates over time and the delivery of oxygen and glucose modulates correspondingly. Experimental descriptions of the coupling between demand and supply were provided over a century ago in laboratory animals (87) and later verified with methods allowing local cerebral flow measurements. Although such methods had been used in conscious laboratory animals since the early 1960s (88), a precise quantitative assessment of the coupling between neural activity and regional blood flow was only possible after the introduction of the deoxyglucose autoradiographic technique that enabled spatially resolved measurements of glucose metabolism in laboratory animals (89). The results of a large number of experiments with the [^{14}C]deoxyglucose method have revealed a clear relationship between local cerebral activation and glucose consumption (90).

The first quantitative measurements of regional brain blood flow and oxygen consumption in humans were performed using the radiotracer techniques, which were followed by the introduction of PET (for a historical review see 91). With PET images, maps of activated brain regions were produced by detecting the indirect

effects of neural activity on variables such as cerebral blood flow (92), cerebral blood volume (CBV) (11), and blood oxygenation (11, 12, 93). At the same time, optical imaging of intrinsic signals demonstrated the precision of neurovascular coupling by constructing detailed maps of cortical microarchitecture in both the anesthetized and the alert animal (94).

Although the existence of a regional coupling between neural activity, metabolism, and hemodynamic changes is now established, the nature of the link between these processes remains an area of study. If energy demand triggers CBF changes, which cellular processes and sites dominate the energy consumption? The structural and functional organization of the neuro-vascular system provides some insights to these questions.

Structural Neurovascular Coupling

One method for studying vascular structure is to infuse low viscosity resins into the vasculature and allow the resin to polymerize. Dissolving away the surrounding tissue with alkali leaves a cast of the three-dimensional distribution of vessels; the cast can then be sectioned and studied with scanning electron microscopy (SEM). Such corrosion cast studies reveal a vast vascular network. These casts also reveal several correlations between vascular density and neuronal structures.

Several investigators have shown that vascular density correlates with the number of synapses, rather than the number of neurons (95–97). For instance, Duvernoy and colleagues demonstrated that the human cortical vascular network could be subdivided based on density into four layers parallel to the surface. These layers systematically overlap with certain portions of the cytoarchitectonically defined Brodmann laminae. The first Duvernoy layer, consisting of vessels oriented approximately parallel to neural fibers, is entirely within the lower part of the molecular layer (Layer I). This layer had the lowest vascularization and IVc had the highest vascularization with a ratio of 3.3:1 (IVc:I, averaged across animals). In layers IVc and I in macaque striate cortex, the synaptic density ratio is 2.43:1, the astrocyte density is 1.2:1, and the neuronal density is 78.8:1. Hence, the vascular density parallels that of the perisynaptic elements rather than the neuronal somata.

A general principle proposed by a number of investigators is that primary sensory areas are characterized by a higher stimulus responsivity and capillary density than association areas and that the vascularization follows cortical plasticity (e.g., 95, 98–103). Figure 5*b* shows a clear example from the Duvernoy et al. study. In this section through human calcarine cortex, one can see a dense vascularization in the input layer (4C) that ends abruptly. The termination of the dense vascularization is in a position that would ordinarily fall near the human V1/V2 boundary.

Similar results were obtained in different cortical areas of rodents. For instance, endovascular casts revealed capillary densities resembling the whisker barrel pattern that characterizes the somatosensory cortex of rats (104) and the spatial patterns of stimulus-induced activation in the auditory cortex of chinchillas (98). It seems that an anatomical neurovascular association exists, and this relationship

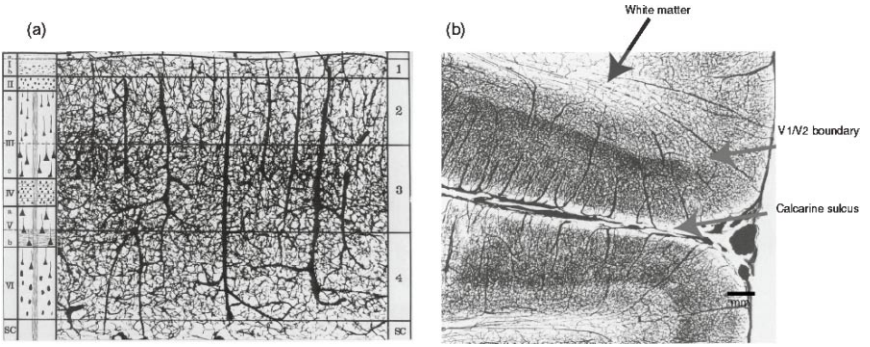

Figure 5 Cortical vascularization across layers and area boundaries. (*a*) Vascularization varies across the cortical layers, with the densest region falling near layers III and IV. Note that layer I is moderately well vascularized. The cell body density in layer I is two orders of magnitude lower than that of layer IVc. The synaptic density is about half that of IVc. Hence, the vascular density parallels synaptic density more closely than cell body density. (*b*) Vascular density varies across boundaries of functional areas. Dense vascularization is present in calcarine cortex, but the density ends abruptly at a location that is likely to be the V1/V2 boundary. These images were adapted from the work of Duvernoy and colleagues (95).

may be a source of the variations in regional coupling between neural activity and metabolism. The pattern of this anatomical neurovascular coupling suggests that blood supply is better correlated with the number of synapses rather than the number of neurons.

Functional Neurovascular Coupling

The cerebral metabolic rate (CMR) is commonly expressed in terms of oxygen consumption ($CMRO_2$). This is a convenient measure because about 90% of the glucose is aerobically metabolized so that the CMR parallels oxygen consumption (105, 106). The energy demands of different neural types probably reflect their electrical activity and physical shape, and size. Large projection neurons, which maintain energy-consuming processes over an extensive membrane surface, may have relatively large average energy requirements. Other cells, including glia and vascular endothelial cells, also play a role in metabolic function, and the specific mechanisms regulating the neurometabolic coupling constitute an area of active investigation. Several interesting cellular and molecular mechanisms and hypotheses have emerged concerning how the cerebral blood flow is coupled to energy consumption.

THE ROLE OF ASTROCYES IN COUPLING There is a tightly regulated glucose metabolism in all brain cell types. An interesting case is the glial cell known as the

astrocyte. The structural and functional characteristics of astrocytes make them ideal bridges between the neuropil and the intraparenchymal capillaries (for detailed references see 107). These specialized glia cells are massively connected with both neurons and the brain's vasculature.

It has been suggested that for each synaptically released glutamate molecule taken up with two to three Na^+ ions by an astrocyte, one glucose molecule enters the same astrocyte, two ATP molecules are produced through glycolysis, and two lactate molecules are released and consumed by neurons to yield 18 ATPs through oxidative phosphorylation. Recent studies established a quantitative relationship between imaging signals and the cycling of certain cerebral neurotransmitters (108–110) because synaptic activity is tightly coupled to glucose uptake (111, 112). Stoichiometric studies using NMR spectroscopy suggest that the utilization of glutamate (Glu), the dominant excitatory neurotransmitter of the brain (about 90% of the synapses in gray matter are excitatory) (113), is equal to the rate at which glutamate is converted to glutamine (Gln) (114). The Glu to Gln conversion occurs in the astrocytes, and the required energy is provided by glycolysis. Astrocytes are indeed enriched in glucose transporters. The transporters are driven by the electrochemical gradient of Na^+; for this reason there is a tight coupling between Glu and Na^+ uptake. Both Glu to Gln conversion and Na^+ restoration require ATP. Gln is subsequently released by astrocytes and taken up by the neuronal terminals to be reconverted to Glu (for review see 13). Calculations based on these findings suggest that the energy demands of glutamatergic neurons account for 80 to 90% of total cortical glucose usage in rats (114) and humans (115).

ENERGY CONSUMPTION Which neural signaling events consume most of the energy? Two groups of investigators calculated an overall energy budget that estimates the energy requirements of neural signaling (16, 120). The budget is based on neural circuitry assumptions concerning the number of vesicles released per action potential, the number of post-synaptic receptors activated per vesicle released, the metabolic consequences of activating a single receptor and changing ion fluxes, and neurotransmitter recycling. The largest portion of energy consumption is attributed to the post-synaptic effects of glutamate (about 34% of the energy in rodents and 74% in humans).

An interesting possibility is that there is a dissociation between the mechanisms calling for the energy and those that use the energy; perhaps neurotransmitter-related signaling, rather than local energy use, drives the hemodynamic responses (121). Indeed, there is evidence that blood flow in a number of brain structures, including neocortex, cerebellum, and hippocampus, may be controlled directly by glutamate and GABA. In the cerebellar cortex, for example, the activation of parallel fibers releases glutamate and leads to the depolarization of Purkinje cells and interneurons. These cells, in turn, release GABA. Notably, the increased blood flow that typically follows the activation of parallel fibers is blocked by inhibitors of non-NMDA glutamate receptors, nitric oxide synthase, and adenosine receptors (122), whereas microinjections of glutamate have vascular effects similar to those

observed during stimulation of the parallel fibers (123). In neocortex and hippocampus, microinjection of neurotransmitters dilates pial arterioles and/or precapillary microvessels, an effect attenuated by inhibitors of nitric oxide synthase (NOS) (124, 125).

These findings are in agreement with microstimulation experiments in which the increase in glucose utilization is assessed during orthodromic and antidromic stimulation. Orthodromic stimulation activates both pre- and post-synaptic terminals, whereas antidromic stimulation activates only post-synaptic terminals. Glucose utilization increased only during orthodromic stimulation (116–118; for review see 119).

Taken together, these results suggest that presynaptic activity (restoration of gradients) and neurotransmitter cycling are the main mechanisms that initiate brain energy production, e.g., via glial pathways. In this view, CBF is controlled by mechanisms that predict energy consumption, not by mechanisms that measure the consumption.

IMPLICATIONS FOR NEURAL CODING While there are uncertainties concerning the mechanisms of CBF control, there is widespread agreement that cortical signaling consumes energy at high rates. This high level of energy consumption may have implications for neural coding strategies because, on the whole, it is better to conserve energy and reduce spike count. But before such a hypothesis can be evaluated, we must also have some measure of the benefits of these signals. The communicating brain, like the beating heart, is doing something worthwhile. A high cost (spike rate) may be worth the benefit for certain circuits and computations.

INTERPRETING THE BOLD SIGNAL

Many investigators use the BOLD response to estimate properties of neural signals. By describing the path from neural signals to the BOLD response, we hope to establish some principles for estimating neural signals from BOLD responses. In this section, we describe a framework to guide such inferences. First, we discuss two qualitative principles, and then we describe formulae and simulations specifically chosen to illustrate the limitations in using the BOLD response to estimate the amplitude and spatial distribution of neural signals.

Qualitative Principles

Given the complexity of the relationship between electrical activity and the BOLD signal, it will be some time before a complete model emerges. However, it is worthwhile to discuss some qualitative principles about the interpretation of the BOLD response. We describe two such principles concerning the interpretation of amplitudes and cortical maps.

COMPARISONS BETWEEN BOLD AND SINGLE-UNIT MEASUREMENTS: AMPLITUDES A number of experiments demonstrate basic differences between the amplitude of electrophysiological and BOLD responses. In one recent study Tolias and colleagues studied brain areas processing motion information (126) using an adaptation technique (127, 128). They repeatedly imaged a monkey's brain while the animal viewed continuous motion in a single, unchanging direction. Under these conditions, the BOLD response adapts. When the direction of motion abruptly reverses, the measured activity immediately shows a partial recovery or rebound. The extent of this rebound is an index of the average directional selectivity in the activated area. The results confirmed previous electrophysiological studies revealing a distributed network of visual areas (V1, V2, V3, V5/MT) in the monkey that process information about the direction of visual motion.

Surprisingly, strong BOLD activation was also observed in area V4; the neuroimaging rebound measurements and thus the inferred motion sensitivity of this area is equal to that of area V5. Yet, single-unit recordings have repeatedly demonstrated very weak motion selectivity in area V4 (e.g., 129).

These results can be interpreted with respect to sub-threshold modulatory signals. Specifically, areas V4 and V5 are extensively interconnected (130–133). Although the principal neurons of these areas may deliver long-range signals that represent different stimulus properties, each area may influence the sensitivity of the other by providing some kind of modulatory input. These short-range inputs may be insufficient to drive the pyramidal cells recorded in a typical electrophysiology experiment. But the BOLD fMRI will measure this modulatory signal and in this way provide measurements that do not match those of single-unit neurophysiology.

Similarly, measurements of visual attention, even measured using the same tasks or stimulation conditions, can differ between fMRI and electrophysiology (134–137; see review 138). A good example is the measurement of the effects of spatial attention on neural activation. Attentional effects on the neurons of area V1 have been very difficult to measure in monkey single-unit experiments (139, 140). Yet using similar tasks strong attentional effects have been readily measurable with fMRI in human V1 (135–137). In addition, attentional effects in area V4 were found to be considerably larger in human fMRI than in monkey electrophysiology (141).

As a general principle, we suggest that some of these differences may arise because the BOLD response emphasizes different aspects of the neural signal: Synaptic activity produced by short-range lateral or feedback input is often more visible with imaging than with single-unit recordings.

COMPARISONS BETWEEN BOLD AND SINGLE-UNIT MEASUREMENTS: MAPS The identification of distinct visual areas, each containing a map-like representation of the visual field, has been an important achievement in neuroscience. A number of groups have developed methods for measuring visual field maps in the human

brain. We have been able to confirm that fMRI methods also measure accurate visual field maps in macaque brain (142). The basic method is to sweep a stimulus from fovea to periphery. As the stimulus travels eccentrically, a wave of activity passes across the visual field representation in cortex. The time at which the activity peaks at each cortical location is an indication of the most effective visual field eccentricity for stimulating that location.

Figure 6 shows an example of the foveal to peripheral representation that can be measured on the operculum (primary visual cortex in macaque). The colors indicate the visual field eccentricity that most strongly drives the fMRI signal at each cortical location.

The graphs in Figure 6 show the time course of the BOLD response in several cortical regions. These regions are each 3-mm radius disks drawn on the surface of the operculum. Although these regions are separated by only a few millimeters, the responses differ markedly, and the time of the peak BOLD response can be easily distinguished. The systematic shift of the time of peak activation reveals the well-known eccentricity map in primary visual cortex and confirms that in this region of the brain the BOLD response and the neural signals are colocalized.

There have been reports that colocalization is poor in somatosensory cortex. Measuring in macaque somatosensory cortex using a 1.5 T magnet and an echo-planar imaging (EPI) pulse sequence, Disbrow and colleagues reported somatosensory maps in areas 3a, 3b, 1, and 2 in separate fMRI and microelectrode recording sessions (143, 144). Although in half the cases the response fields overlapped, in the other half the fMRI signal was displaced from the electrical signal, sometimes by as much as a centimeter. Those measurements suggest that the ability to colocalize may turn out to depend on factors including the position within the cortical sheet, perhaps coupled to differences in the vascularization patterns, or to differences in experimental protocols (145).

Hemodynamic Response Efficiency

As we have reviewed here, the coupling between neural activity and the vascular response is significant in determining the amplitude and spatial resolution of the BOLD signal. We refer to the efficacy of this coupling as the hemodynamic response efficiency (HRE). Regions of sparse vascularization are likely to have low HRE and weak or absent BOLD response.

It is conceivable that below a certain vascularization density hemodynamic responses cannot be detected despite increases in neural activity and energy metabolism. White matter is a good example. The energy consumption in white matter is one fourth that of gray matter. This consumption has been measured in studies of glucose-metabolism and of blood flow changes using intracerebral O^{15}-H_2O microprobes (B. Weber, personal communication). White matter consumes energy for the restoration of ionic gradients perturbed by the spreading action potential at the Ranvier nodes where the Na^+/K^+-dependent ATPase is mainly responsible for the increased metabolic demands of brain tissue (146). In

addition, neurotransmitter-mediated signaling may take place along fiber tracts lacking vesicular means of releasing neurotransmitters (for review see 147). In this receptor-mediated signaling mechanism, glutamate is released from axons via the reversal of a transporter to induce intracellular calcium spiking in glial cells by means of metabotropic glutamate receptors. The mechanism may serve the axon-glia interactions that are most likely involved in glutamate-induced glycolysis, such as that described for astrocytes in the vicinity of glutamatergic synapses (see for example 13). Yet, activation of the white matter has been rarely reported in the neuroimaging literature (148, 149), and a reasonable investigator may doubt the presence of a BOLD signal in white matter altogether.

An additional constraint for spatial resolution comes from the organization of the parenchymal blood supply. Pial arteries, often arising from arterial trunks, as well as capillary branching points very often have a slight rosary-like constriction at their point of origin (95, 98, 150, 151). These constrictions are widely believed to indicate the presence of muscular sphincters regulating the cortical arterial flow. Veins were never reported to show similar changes of diameter. We hypothesize that the density of such myogenic valves will partly determine the HRE and thus whether a cortical region exhibits powerful BOLD responses.

BOLD response measurements also may be influenced by precapillary shunts. These are anastomoses or arterio-venous shunts that serve as alternative channels and have been demonstrated in different tissues (152). When metabolic needs are low, the sphincters of the arterioles close and blood bypasses the capillary bed, flowing through the alternative channels directly to the venules. The pressure elevation that follows the increased metabolic demands opens the sphincters, and the blood flows again through the capillary bed. Precapillary arterio-venous shunts were rarely found in cortex (153) and were entirely absent in the corrosion cast studies (95). On the other hand, the vascular dilatation responsible for the increase in blood flow evoked by neural activity has been shown to propagate in a retrograde fashion to upstream arterioles located outside the activated area (154); a process quite possibly interfering with the spatial specificity of high-resolution imaging.

Finally, simultaneous recording of intrinsic signals and measurements of parenchymal flow with H_2 electrodes showed that activation increasing flow to a particular whisker barrel often leads to reduced flow to adjacent cortex (153). Such a vascular-contrast mechanism enhances local differences and may favorably influence the spatial resolution of a neuroimaging method. Caution is required here, too, because areas of activation may not always match the tissue region from which the signals are generated (98).

All in all, the anatomical substrate of the hemodynamic responses suggests a need for extreme caution when precise localization of function is required and quantitative comparisons must be made between signals coming from different areas. The HRE of each brain region may differ from other regions, even when the regions are nearby. It seems that with our current knowledge there is no secure way to determine a quantitative relationship between a hemodynamic response amplitude and its underlying neural activity in terms of either number of spikes

per unit time per BOLD increase or amount of perisynaptic activity. This holds true even if BOLD is calibrated by taking into account the local perfusion of the tissue (156, 157).

The BOLD Signal Measures Circuit Properties

The data summarized above suggest that synaptic potentials are the strongest cause of the BOLD responses. An important consequence of this hypothesis is that the BOLD response will differ depending on the neural circuit properties within various regions of cortex.

Consider the implications in the case of two hypothetical neural circuits located in different regions of cortex. We suppose that in both regions a specific stimulus causes, on average, 10 action potentials above baseline. The two cortical regions differ, however, in that the first cortical region contains a circuit that measures the size of a purely excitatory input, whereas the circuitry in the second region measures the difference between opposing excitatory and inhibitory inputs. On the hypothesis that the BOLD response depends on synaptic potentials, the purely excitatory circuit usually will generate a smaller BOLD response than the circuit that compares opposing inputs. Hence, other factors being equal, the BOLD signal in the second region should exceed that in the first, despite the equal spiking activity.

In this example the amplitude of the BOLD response per action potential will differ across the two regions. Because we have assumed that both regions produce the same number of spikes, the region with the direct excitatory circuitry will have a smaller BOLD response/spike than the region with opposing inputs. Circuitry, then, is an additional reason for differential coupling between action potentials and BOLD responses.

A corollary of this reasoning is that the relationship between action potentials and BOLD responses contains useful information about local circuitry. For example, estimates of the number of spikes per 1% BOLD response in V1 and MT differ by an order of magnitude (80, 158; for a review see 159). If the incremental spikes in MT are produced by a direct excitatory connection, the activity associated with nine spikes might be enough to produce 1% BOLD response. If V1 signals are based on the comparisons of multiple signals, then a great deal of dendritic activity may be required to produce a small number of spikes. This hypothesis can be tested by an experimental analysis of the circuitry in these regions.

A Mathematical Framework

To increase the precision of the inferences we can make from BOLD responses to neural signals, it is useful to summarize the relationship in some simple equations and figures. In the following, we simulate the relationship between simple patterns of neural activity and the BOLD signal. We offer this simulation as a general guide to help reason about the relationship between the BOLD and neural signals.

BOLD Amplitude

We refer to the spatial pattern of stimulus-driven neural activity during an experiment as $A(x)$, and we describe the resulting spatial pattern of BOLD responses as $B(x)$. The spatial variable, x, refers to the position along the cortical surface. Our goal is to understand the relationship between these two quantities.

In addition to stimulus-driven activity, there will be uncontrolled neural responses that we summarize as a random variable, $\underset{\sim}{N}_N(x)$. These responses represent the ongoing neural noise. (We use the tilda to indicate that the term is a random variable). Hence, the total neural activity is $A(x) + \underset{\sim}{N}_N(x)$. An example of one pattern of neural activity and noise is shown in Figure 7*a*.

At each location on the cortical surface, the neural activity is more or less efficiently coupled to the vascular response by the HRE. We express this efficiency as $H(x)$. The spatial distribution of the HRE is shown in Figure 7*b*. The vascular demand imposed at each point is $(A(x) + \underset{\sim}{N}_N(x))H(x)$.

The neural demand on the vasculature spreads across the cortical surface so that the BOLD signal is a blurred and noisy representation of this vascular demand. The degree of blurring, say on the order of 5 mm (full-width, half maximum) in primary visual cortex (160); of course, this quantity may not generalize across cortex. We use this number to illustrate the key issues for the moment. Mathematically, we can approximate the spatial spreading at x as the weighted sum of the demand in a nearby neighborhood, $n(x)$, on the cortical surface. The spread is measured by the pointspread function, $P(x)$.

$$\int_{n(x)} (A(u) + \underset{\sim}{N}_N(u))H(u)P(x-u)\,du.$$

Finally, as in all measurement systems, there is measurement noise, $\underset{\sim}{N}_M$. In this case, the noise can be attributed to factors such as the magnetic resonance instrument, brain pulsatility, and other uncontrollable experimental factors. We can express the relationship between the neural activity, $A(x)$, and the BOLD response, $B(x)$, by the equation

$$B(x) = \int_{n(x)} (A(u) + \underset{\sim}{N}_N(u))H(u)P(x-u)\,du + \underset{\sim}{N}_M(x).$$

The pseudo-color image in Figure 7*c* shows the measured BOLD response to the pattern of neural activity in Figure 7*a*.

The conventional subtraction methodology measures a BOLD response as the difference between an active stimulus condition and a control condition. Suppose that the stimulus-driven neural activity in the control condition is $\underline{A}$. The BOLD measurement described in most fMRI experiments is a spatial map of this difference.

$$\begin{aligned}\Delta B(x) &= B(x) - \bar{B}(x) \\ &= \int_{n(x)} (A(u) + \underset{\sim}{N}_N(u))H(u)P(x-u)\,du + \underset{\sim}{N}_M(x) \\ &\quad - \int_{n(x)} (\bar{A}(u) + \underset{\sim}{N}_N(u))H(u)P(x-u)\,du + \underset{\sim}{N}_M(x) \\ &= \int_{n(x)} (A(u) - \bar{A}(u) + 2\underset{\sim}{N}_N(u))H(u)P(x-u)\,du + 2\underset{\sim}{N}_M(x).\end{aligned}$$

Figure 8*a* shows the differential BOLD response in a simulation that includes all types of noise. Figure 8*b* shows the locations where the differential BOLD response exceeds a relatively high threshold.

Consider a few basic features of the differential BOLD response formula. First, the expected values of the noise terms are zero (the factor of two arises because the two noise terms are subtracted). Second, the expected amplitude of the differential BOLD response is a spatially blurred version of the stimulus-driven differential activity multiplied by the HRE, $(A(x) - \bar{A}(x))H(x)$. It is only the stimulus-driven difference that is of experimental interest, but this difference cannot be directly estimated because the HRE is unknown. Consequently, a direct comparison of the amplitude of BOLD responses between well-separated cortical regions should only be made if one has some reason to expect that the HRE values in these two regions are equal, or at least known. It may be reasonable to compare the responses at nearby locations, assuming that the HRE varies smoothly across cortex or when we have other data about the HRE.

Even some simple conclusions are problematic when we consider this framework. For example, when the BOLD measurements at two locations are ordered $B(x) > B(y)$, without some information about the HRE, we cannot conclude securely that $A(x) > A(y)$.

BOLD Signal-to-Noise Ratio

Because of the heavy reliance on statistical reasoning in neuroimaging, many groups report a statistical map of the activity based on a variety of measures (z-score, p-value, etc.). In these maps, the statistical parameter represents the amplitude of the differential BOLD measurement relative to noise.

While there are many statistical models, and hence statistical parameter maps, all such models are based on some type of comparison of the size of the differential BOLD response to noise. The values in these parameter maps can be summarized by the signal-to-noise ratio, in which the signal term is in the numerator and the noise terms in the denominator,

$$SNR(x) \sim \frac{\int_{n(x)} (A(u) - \bar{A}(u))H(u)P(x-u)\,du}{2\underset{\sim}{N}_M + 2\int_{n(x)} H(u)\underset{\sim}{N}_N(u)P(x-u)\,du}.$$

In general, high signal-to-noise ratios are associated with a large z-score (smaller p-values). The precise relationship between these quantities depends upon the specific statistical model assumptions. Examining the SNR equation, we note a couple of simple properties.

First, suppose that (*a*) the measurement noise $\underset{\sim}{N}_M$ is small, and (*b*) the spatial spread, P(), is very localized. Then the signal-to-noise ratio simplifies to

$$SNR \sim \frac{(A - \bar{A})}{2\underset{\sim}{N}_N}.$$

This SNR model is implicit in many neuroimaging investigations; the hemodynamic response efficiency, H, disappears and the SNR is proportional to the signal difference divided by the noise.

Alternatively, consider the case in which (*a*) the instrumental noise dominates, and (*b*) the spatial spread is again modest. In that case the signal-to-noise ratio becomes

$$SNR \sim \frac{(A - \bar{A})H}{2\underset{\sim}{N}_M}.$$

In this case, the SNR measures the signal difference combined with the HRE.

In general, we do not know which noise dominates. In fact, the operating regime may depend on features of the instrument, pulse-sequence selection, voxel size, and so forth in complex ways.

The main purpose of working through these formulae is to alert us to some simple principles that should be kept in mind when inferring neural signals from BOLD responses. The simulation parameters were chosen to illustrate the effects of the HRE and spatial blurring of the differential BOLD response. In particular, the spatial periodicity of the neural activity in one direction (dorsal to ventral) was chosen to be too fine to be captured by the blurred BOLD response. The spatial variation in the anterior to posterior direction, however, is of sufficiently low spatial frequency that it passes through the blurring. Portions of this activity are unseen because the simulation uses a low HRE value on the dorsal surface. Hence, with these parameters the BOLD response simulation fails to mirror several aspects of the neural activity.

By choosing these parameters, which are possible but not necessarily true, we hope to encourage further research into measurements of the HRE and spatial blurring. Given the many successes of BOLD neuroimaging, the simulation parameters we have chosen may be too pessimistic. However, it would be far better to know the values of these parameters from direct measurements in specific regions of the

brain. We hope, therefore, that this review and simulations will serve to motivate further investigation of these quantities.

CONCLUSIONS

In the short period of time since its introduction, fMRI has evolved to become the most important method for investigating human brain function. The BOLD response provides us unprecedented visibility of the neural activity in the human brain; a visibility that far exceeds previously available methods. Still, the visibility remains limited, and the data must be interpreted with these limitations in mind.

A number of important issues remain unresolved, and these issues require careful attention if we want to interpret neural mechanisms from this surrogate signal. The BOLD contrast mechanism reflects changes in cerebral blood volume, cerebral blood flow, and oxygen consumption. The interaction between neural activity and these variables involves a number of factors, including the cell types and circuitry driven during activation, and the processes that couple energy demand to its supply to the brain (HRE).

Research in a number of fields ranging from biochemistry, biophysics, and molecular biology to physics and engineering provides us daily with new information regarding cellular events that may be involved in the generation of responses or the technology that may be best applied to understand these aspects of the fMRI signals.

The first simultaneous fMRI and electrophysiological recordings clearly confirmed a longstanding assumption, i.e., the BOLD contrast mechanism reflects aspects of the neural responses elicited by a stimulus. The hemodynamic response primarily reflects the neuronal input to the relevant area of the brain and its processing there rather than the long-range signals transmitted by action potentials to other regions of the brain. It is reasonable to expect that output activity will usually correlate with neurotransmitter release and pre- and post-synaptic currents. But when input into a particular area plays a primarily modulatory role, fMRI experiments may measure activation that does not correlate well with single-unit measurements.

From examining the pathway between neural signals and BOLD response, we conclude that comparisons of the response amplitude at two locations may not reflect the neural signal amplitudes at these locations. Measuring (*a*) the relative tuning to various stimuli at a single cortical location and (*b*) spatial maps yields more secure conclusions. The relative tuning at a single location may be based on a single coupling, although even this is not certain because of laminar differences. And when one observes a map, it is likely to represent a neural map. Of course, the inability to measure a map using BOLD does not imply the absence of a neural map.

Finally, a great deal of research will be needed to elucidate the neurometabolic coupling and its subsequent hemodynamic response. Neuronal synaptic activity responses may be specific for stimuli and brain areas and are frequently nonlinearly related to the stimulus properties. Although glucose consumption is known to

increase during the activity of inhibitory synaptic activity, the role of inhibition in the generation of BOLD remains elusive.

Despite these and other questions, fMRI has earned an indispensable position in neuroscience and, if combined with other invasive techniques, it promises a much better understanding of information processing in the central nervous system than any other technology alone.

ACKNOWLEDGMENTS

We thank Alyssa Brewer for reading and commenting on the manuscript and help with figure preparation. We thank Robert Dougherty, Edgar Galindo, Rainer Goebel, Peter Lennie, Marcus Raichle, David Ress, and Whitman Richards for their comments and suggestions, and Daniel Donoho for help in manuscript preparation. Supported by the Max-Planck Society and NEI RO1 EY03164.

The *Annual Review of Physiology* is online at http://physiol.annualreviews.org

LITERATURE CITED

1. Callaghan PT. 1991. *Principles of Nuclear Magnetic Resonance Microscopy.* New York: Oxford Univ. Press
2. Schmitt F, Stehling MK, Turner R. 1998. *Echo-Planar Imaging: Theory, Technique and Application.* Berlin: Springer
3. Pauling L, Coryell C. 1936. The magnetic properties and structure of hemoglobin. *Proc. Natl. Acad. Sci. USA* 22:210–16
4. Brooks RA, Battocletti JH, Sances A, Larson SJ, Bowman RL, Kudravcev V. 1975. Nuclear magnetic relaxation in blood. *IEEE Trans. Biomed. Eng.* 22:12–18
5. Thulborn KR, Waterton JC, Matthews PM, Radda GK. 1982. Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. *Biochim. Biophys. Acta* 714:265–70
6. Thulborn KR, Waterton JC, Matthews PM. 1992. Dependence of the transverse relaxation time of water protons in whole blood at high field. *Biochem. Biophys. Acta* 714:265–72
7. Ogawa S, Lee TM, Nayak AS, Glynn P. 1990. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn. Reson. Med.* 14:68–78
8. Ogawa S, Lee TM, Kay AR, Tank DW. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. USA* 87:9868–72
9. Ogawa S, Lee TM. 1990. Magnetic resonance imaging of blood vessels at high fields: in vivo and in vitro measurements and image simulation. *Magn. Reson. Med.* 16:9–18
10. Turner R, Le Bihan D, Moonen CT, Des-Pres D, Frank J. 1991. Echo-planar time course MRI of cat brain oxygenation changes. *Magn. Reson. Med.* 22:159–66
11. Fox PT, Raichle ME. 1986. Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proc. Natl. Acad. Sci. USA* 83:1140–44
12. Fox PT, Raichle ME, Mintun MA, Dence C. 1988. Nonoxidative glucose consumption during focal physiologic neural activity. *Science* 241:462–64

13. Magistretti PJ, Pellerin L. 1999. Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging. *Philos. Trans. R. Soc. London Ser. B.* 354:1155–63
14. Hyder F, Shulman RG, Rothman DL. 1998. A model for the regulation of cerebral oxygen delivery. *J. Appl. Physiol.* 85:554–64
15. Buxton RB, Frank LR. 1997. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *J. Cereb. Blood Flow Metab.* 17:64–72
16. Attwell D, Laughlin SB. 2001. An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* 21:1133–45
17. Mintun MA, Lundstrom BN, Snyder AZ, Vlassenko AG, Shulman GL, Raichle ME. 2001. Blood flow and oxygen delivery to human brain during functional activity: theoretical modeling and experimental data. *Proc. Natl. Acad. Sci. USA* 98:6859–64
18. Shimojyo S, Scheinberg P, Kogure K, Reinmuth OM. 1968. The effect of graded hypoxia upon transient cerebral blood flow and oxygen consumption. *Neurology* 18:127–33
19. Glover GH. 1999. Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage* 9:416–29
20. Martindale J, Mayhew J, Berwick J, Jones M, Martin C, et al. 2003. The hemodynamic impulse response to a single neural event. *J. Cereb. Blood Flow Metab.* 23:546–55
21. Yang Y, Engelien W, Pan H, Xu S, Silbersweig DA, Stern E. 2000. A CBF-based event-related brain activation paradigm: characterization of impulse-response function and comparison to BOLD. *Neuroimage* 12:287–97
22. Pu Y, Liu HL, Spinks JA, Mahankali S, Xiong J, et al. 2001. Cerebral hemodynamic response in Chinese (first) and English (second) language processing revealed by event-related functional MRI. *Magn. Reson. Imaging* 19:643–47
23. Liu H, Gao JH. 2000. An investigation of the impulse functions for the nonlinear BOLD response in functional MRI. *Magn. Reson. Imaging* 18:931–38
24. Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, et al. 1992. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl. Acad. Sci. USA* 89:5675–79
25. Buxton RB, Wong EC, Frank LR. 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn. Reson. Med.* 39:855–64
26. Frahm J, Kruger G, Merboldt KD, Kleinschmidt A. 1996. Dynamic uncoupling and recoupling of perfusion and oxidative metabolism during focal brain activation in man. *Magn. Reson. Med.* 35:143–48
27. Logothetis NK, Guggenberger H, Peled S, Pauls J. 1999. Functional imaging of the monkey brain. *Nat. Neurosci.* 2:555–62
28. Robson MD, Dorosz JL, Gore JC. 1998. Measurements of the temporal fMRI response of the human auditory cortex to trains of tones. *Neuroimage* 7:185–98. Erratum. *Neuroimage.* 1998. 8:228
29. Logothetis NK. 2002. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philos. Trans. R. Soc. London Ser. B* 357:1003–37
30. Boynton GM, Engel SA, Glover GH, Heeger DJ. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* 16:4207–21
31. Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R. 1998. Event-related fMRI: characterizing differential responses. *Neuroimage* 7:30–40
32. Worsley KJ. 2002. Statistical analysis of activation images. In *Functional Magnetic Resonance Imaging: An Introduction to Methods*, ed. P Jezzard, PM

Matthews, SM Smith, pp. 251–70. York: Oxford Univ. Press
33. Bandettini PA, Wong EC, Hinks RS, Tikofsky RS, Hyde JS. 1992. Time course EPI of human brain function during task activation. *Magn. Reson. Med.* 25:390–97
34. Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, et al. 1992. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA* 89:5951–55
35. Engel SA, Glover GH, Wandell BA. 1997. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex* 7:181–92
36. Ranck JBJ. 1963. Specific impedance of rabbit cerebral cortex. *Exp. Neurol.* 7:144–52
37. Ranck JBJ. 1966. Electrical impedance in the subicular area of rats during paradoxical sleep. *Exp. Neurol.* 16:416–37
38. Mitzdorf U. 1985. Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiol. Rev.* 65:37–100
39. Nicholson C, Freeman JA. 1975. Theory of current source-density analysis and determination of conductivity tensor for anuran cerebellum. *J. Neurophysiol.* 38:356–68
40. Gray CM, Maldonado PE, Wilson M, McNaughton B. 1995. Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *J. Neurosci. Methods* 63:43–54
41. Harris KD, Henze DA, Csicsvari J, Hirase H, Buzsaki G. 2000. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J. Neurophysiol.* 84:401–14
42. Henze DA, Borhegyi Z, Csicsvari J, Mamiya A, Harris KD, Buzsaki G. 2000. Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *J. Neurophysiol.* 84:390–400
43. Adrian ED, Zotterman Y. 1926. The impulses produced by sensory nerve-endings, Part 2. The response of a single end-organ. *J. Physiol.* 61:151–71
44. Kuffler SW. 1953. Discharge patterns and functional organization of the mammalian retina. *J. Neurophysiol.* 16:37–68
45. Hubel DH, Wiesel TN. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160:106–54
46. Mountcastle VB. 1957. Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J. Neurophysiol.* 20:408–34
47. Lettvin JY, Maturana HR, McCulloch WS, Pitts WH. 1959. What the frog's eye tells the frog's brain. *Proc. Inst. Radio Engrs.* 47:1940–51
48. Stone J. 1973. Sampling properties of microelectrodes assessed in the cat's retina. *J. Neurophysiol.* 36:1071–79
49. Towe AL, Harding GW. 1970. Extracellular microelectrode sampling bias. *Exp. Neurol.* 29:366–81
50. Rall W. 1962. Electrophysiology of a dendritic neuron. *Biophys. J.* 2:145–67
51. Humphrey DR, Corrie WS. 1978. Properties of pyramidal tract neuron system within a functionally defined subregion of primate motor cortex. *J. Neurophysiol.* 41:216–43
52. Bishop GH, O'Leary JL. 1942. Factors determining the form of the potential record in the vicinity of the synapses of the dorsal nucleus of the lateral geniculate body. *J. Cell Comp. Physiol.* 19:315–31
53. Lorente de Nó R. 1947. Action potentials of the motorneurones of the hypoglossus nucleus. *J. Cell Comp. Physiol.* 29:207–88
54. Logothetis NK. 2003. The underpinnings of the BOLD functional magnetic resonance imaging signal. *J. Neurosci.* 23:3963–71
55. Logothetis NK. 2002. On the neural basis

of the BOLD fMRI signal. *Philos. Trans. R. Soc. London Ser. B.* 357:1003–37
56. Buchwald JS, Grover FS. 1970. Amplitudes of background fast activity characteristic of specific brain sites. *J. Neurophysiol.* 33:148–59
57. Nelson PG. 1966. Interaction between spinal motoneurons of the cat. *J. Neurophysiol.* 29:275–87
58. Grover FS, Buchwald JS. 1970. Correlation of cell size with amplitude of background fast activity in specific brain nuclei. *J. Neurophysiol.* 33:160–71
59. Fromm GH, Bond HW. 1964. Slow changes in the electrocorticogram and the activity of cortical neurons. *Electroencephalogr. Clin. Neurophysiol.* 17:520–23
60. Fromm GH, Bond HW. 1967. The relationship between neuron activity and cortical steady potentials. *Electroencephalogr. Clin. Neurophysiol.* 22:159–66
61. Ajmone-Marsan C. 1965. Electrical activity of the brain: slow waves and neuronal activity. *Isr. J. Med. Sci.* 1:104–17
62. Buchwald JS, Hala ES, Schramm S. 1965. A comparison of multi-unit activity EEG activity recorded from the same brain site in chronic cats during behavioral conditioning. *Nature* 205:1012–14
63. Mitzdorf U. 1987. Properties of the evoked potential generators: current source-density analysis of visually evoked potentials in the cat cortex. *Int. J. Neurosci.* 33:33–59
64. Juergens E, Guettler A, Eckhorn R. 1999. Visual stimulation elicits locked and induced gamma oscillations in monkey intracortical- and EEG-potentials, but not in human EEG. *Exp. Brain Res.* 129:247–59
65. Kamondi A, Acsady L, Wang XJ, Buzsaki G. 1998. Theta oscillations in somata and dendrites of hippocampal pyramidal cells in vivo: activity-dependent phase-precession of action potentials. *Hippocampus* 8:244–61
66. Granit R, Kernell D, Smith RS. 1963. Delayed depolarization and the repetitive response to intracellular stimulation of mammalian motoneurones. *J. Physiol.* 168:890–910
67. Harada Y, Takahashi T. 1983. The calcium component of the action potential in spinal motoneurones of the rat. *J. Physiol.* 335:89–100
68. Gustafsson B. 1984. Afterpotentials and transduction properties in different types of central neurones. *Arch. Ital. Biol.* 122:17–30
69. Chandler SH, Hsaio CF, Inoue T, Goldberg LJ. 1994. Electrophysiological properties of guinea pig trigeminal motoneurons recorded in vitro. *J. Neurophysiol.* 71:129–45
70. Walton K, Fulton BP. 1986. Ionic mechanisms underlying the firing properties of rat neonatal motoneurons studied in vitro. *Neuroscience* 19:669–83
71. Higashi H, Tanaka E, Inokuchi H, Nishi S. 1993. Ionic mechanisms underlying the depolarizing and hyperpolarizing afterpotentials of single spike in guinea-pig cingulate cortical neurons. *Neuroscience* 55:129–38
72. Kobayashi M, Inoue T, Matsuo R, Masuda Y, Hidaka O, et al. 1997. Role of calcium conductances on spike afterpotentials in rat trigeminal motoneurons. *J. Neurophysiol.* 77:3273–83
73. Buzsaki G. 1931. Theta oscillations in the hippocampus. *Neuron* 33:325–40
74. Buzsaki G, Bickford RG, Ponomareff G, Thal LJ, Mandel R, Gage FH. 1988. Nucleus basalis and thalamic control of neocortical activity in the freely moving rat. *J. Neurosci.* 8:4007–26
75. Bullock TH. 1997. Signals and signs in the nervous system: the dynamic anatomy of electrical activity is probably information-rich. *Proc. Natl. Acad. Sci. USA* 94:1–6
76. Wandell BA. 1999. Computational neuroimaging of human visual cortex. *Annu. Rev. Neurosci.* 22:145–73
77. Brinker G, Bock C, Busch E, Krep H, Hossmann KA, Hoehn-Berlage M. 1999.

Simultaneous recording of evoked potentials T2*-weighted MR images during somatosensory stimulation of rat. *Magn. Reson. Med.* 41:469–73
78. Mathiesen C, Caesar K, Akgoren N, Lauritzen M. 1998. Modification of activity-dependent increases of cerebral blood flow by excitatory synaptic activity and spikes in rat cerebellar cortex. *J. Physiol.* 512:555–66
79. Ogawa S, Lee TM, Stepnoski R, Chen W, Zhu XH, Ugurbil K. 2000. An approach to probe some neural systems interaction by functional MRI at neural time scale down to milliseconds. *Proc. Natl. Acad. Sci. USA* 97:11026–31
80. Rees G, Friston K, Koch C. 2000. A direct quantitative relationship between the functional properties of human and macaque V5. *Nat. Neurosci.* 3:716–23
81. Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150–57
82. Hyder F, Rothman DL, Shulman RG. 2002. Total neuroenergetics support localized brain activity: implications for the interpretation of fMRI. *Proc. Natl. Acad. Sci. USA* 99:10771–76
83. Smith AJ, Blumenfeld H, Behar KL, Rothman DL, Shulman RG, Hyder F. 2002. Cerebral energetics and spiking frequency: The neurophysiological basis of fMRI. *Proc. Natl. Acad. Sci. USA* 99:10765–70
84. Lauritzen M, Gold L. 2003. Brain function and neurophysiological correlates of signals used in functional neuroimaging. *J. Neurosci.* 23:3972–80
85. Lauritzen M. 2001. Relationship of spikes, synaptic activity, and local changes of cerebral blood flow. *J. Cereb. Blood Flow Metab.* 21:1367–83
86. Mathiesen C, Caesar K, Lauritzen M. 2000. Temporal coupling between neuronal activity and blood flow in rat cerebellar cortex as indicated by field potential analysis. *J. Physiol.* 523:235–46
87. Roy CS, Sherrington CS. 1890. On the regulation of the blood supply of the brain. *J. Physiol.* 11:85–108
88. Sokoloff L. 1981. Relationships among local functional activity, energy metabolism, and blood flow in the central nervous system. *Fed. Proc.* 40:2311–16
89. Sokoloff L, Reivich M, Kennedy C, DesRosiers MH, Patlak CS, et al. 1977. The ^{C14}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
90. Sokoloff L. 1977. Relation between physiological function and energy metabolism in the central nervous system. *J. Neurochem.* 29:13–26
91. Raichle ME. 2000. A brief history of human functional brain mapping. In *The Systems*, ed. AW Toga, JC Mazziotta, pp. 33–75. San Diego: Academic
92. Fox PT, Mintun MA, Raichle ME, Miezin FM, Allman JM, Van Essen DC. 1986. Mapping human visual cortex with positron emission tomography. *Nature* 323:806–9
93. Frostig RD, Lieke EE, Ts'o DY, Grinvald A. 1990. Cortical functional architecture and local coupling between neuronal activity and the microcirculation revealed by in vivo high-resolution optical imaging of intrinsic signals. *Proc. Natl. Acad. Sci. USA* 87:6082–86
94. Bonhoeffer T, Grinvald A. 1996. Optical imging based on intrinsic signals. In *Brain Mapping, The Methods*, ed. AW Toga, JC Mazziotta, pp. 55–97. New York: Academic
95. Duvernoy HM, Delon S, Vannson JL. 1981. Cortical blood vessels of the human brain. *Brain Res. Bull.* 7:519–79
96. Schuz A, Palm G. 1989. Density of neurons and synapses in the cerebral cortex of the mouse. *J. Comp. Neurol.* 286:442–55
97. O'Kusky J, Colonnier M. 1982. A laminar analysis of the number of neurons,

glia, and synapses in the adult cortex (area 17) of adult macaque monkeys. *J. Comp. Neurol.* 210:278–90

98. Harrison RV, Harel N, Panesar J, Mount RJ. 2002. Blood capillary distribution correlates with hemodynamic-based functional imaging in cerebral cortex. *Cereb. Cortex* 12:225–33
99. Swain RA, Harris AB, Wiener EC, Dutka MV, Morris HD, et al. 2003. Prolonged exercise induces angiogenesis and increases cerebral blood volume in primary motor cortex of the rat. *Neuroscience* 117:1037–46
100. Argandona EG, Rossi ML, Lafuente JV. 2003. Visual deprivation effects on the s 100beta positive astrocytic population in the developing rat visual cortex: a quantitative study. *Brain Res. Dev. Brain Res.* 141:63–69
101. Argandona EG, Lafuente JV. 2000. Influence of visual experience deprivation on the postnatal development of the microvascular bed in layer IV of the rat visual cortex. *Brain Res.* 855:137–42
102. Black JE, Isaacs KR, Anderson BJ, Alcantara AA, Greenough WT. 1990. Learning causes synaptogenesis, whereas motor activity causes angiogenesis, in cerebellar cortex of adult rats. *Proc. Natl. Acad. Sci. USA* 87:5568–72
103. Toga AW. 1987. The metabolic consequence of visual deprivation in the rat. *Brain Res.* 465:209–17
104. Cox SB, Woolsey TA, Rovainen CM. 1993. Localized dynamic changes in cortical blood flow with whisker stimulation corresponds to matched vascular and neuronal architecture of rat barrels. *J. Cereb. Blood Flow Metab.* 13:899–913
105. Siesjo BoK. 1978. *Brain Energy Metabolism.* New York: Wiley & Sons
106. Ames A. 2000. CNS energy metabolism as related to function. *Brain Res. Brain Res. Rev.* 34:42–68
107. Araque A, Parpura V, Sanzgiri RP, Haydon PG. 1999. Tripartite synapses: glia, the unacknowledged partner. *Trends Neurosci.* 22:208–15
108. Magistretti PJ, Pellerin L, Rothman DL, Shulman RG. 1999. Neuroscience—energy on demand. *Science* 283:496–97
109. Rothman DL, Sibson NR, Hyder F, Shen J, Behar KL, Shulman RG. 1999. In vivo nuclear magnetic resonance spectroscopy studies of the relationship between the glutamate-glutamine neurotransmitter cycle and functional neuroenergetics. *Philos. Trans. R. Soc. London Ser. B* 354:1165–77
110. Shulman RG, Rothman DL. 1998. Interpreting functional imaging studies in terms of neurotransmitter cycling. *Proc. Natl. Acad. Sci. USA* 95:11993–98
111. Takahashi S, Driscoll BF, Law MJ, Sokoloff L. 1995. Role of sodium and potassium ions in regulation of glucose metabolism in cultured astroglia. *Proc. Natl. Acad. Sci. USA* 92:4616–20
112. Pellerin L, Magistretti PJ. 1994. Glutamate uptake into astrocytes stimulates aerobic glycolysis: a mechanism coupling neuronal activity to glucose utilization. *Proc. Natl. Acad. Sci. USA* 91:10625–29
113. Braitenberg V, Schuez A. 1998. *Cortex: Statistics and Geometry of Neuronal Connectivity.* Berlin: Springer
114. Sibson NR, Dhankhar A, Mason GF, Rothman DL, Behar KL, Shulman RG. 1998. Stoichiometric coupling of brain glucose metabolism and glutamatergic neuronal activity. *Proc. Natl. Acad. Sci. USA* 95:316–21
115. Pan JW, Stein DT, Telang F, Lee JH, Shen J, et al. 2000. Spectroscopic imaging of glutamate C4 turnover in human brain. *Magn. Reson. Med.* 44:673–79
116. Kadekaro M, Crane AM, Sokoloff L. 1985. Differential effects of electrical stimulation of sciatic nerve on metabolic activity in spinal cord and dorsal root ganglion in the rat. *Proc. Natl. Acad. Sci. USA* 82:6010–13
117. Kadekaro M, Vance WH, Terrell ML,

Gary HJ, Eisenberg HM, Sokoloff L. 1987. Effects of antidromic stimulation of the ventral root on glucose utilization in the ventral horn of the spinal cord in the rat. *Proc. Natl. Acad. Sci. USA* 84:5492–95

118. Nudo RJ, Masterton RB. 1986. Stimulation-induced ^{14}C2-deoxyglucose labeling of synaptic activity in the central auditory system. *J. Comp. Neurol.* 245:553–65
119. Jueptner M, Weiller C. 1995. Review—does measurement of regional cerebral blood flow reflect synaptic activity—implications for PET and fMRI. *Neuroimage* 2:148–56
120. Lennie P. 2003. The cost of cortical computation. *Curr. Biol.* 13:493–97
121. Attwell D, Iadecola C. 2002. The neural basis of functional brain imaging signals. *Trends Neurosci.* 25:621–25
122. Li J, Iadecola C. 1994. Nitric oxide and adenosine mediate vasodilation during functional activation in cerebellar cortex. *Neuropharmacology* 33:1453–61
123. Yang G, Iadecola C. 1996. Glutamate microinjections in cerebellar cortex reproduce cerebrovascular effects of parallel fiber stimulation. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 271:R1568–75
124. Faraci FM, Breese KR. 1993. Nitric oxide mediates vasodilatation in response to activation of *N*-methyl-D-aspartate receptors in brain. *Circ. Res.* 72:476–80
125. Fergus A, Lee KS. 1997. Regulation of cerebral microvessels by glutamatergic mechanisms. *Brain Res.* 754:35–45
126. Tolias AS, Smirnakis SM, Augath MA, Trinath T, Logothetis NK. 2001. Motion processing in the macaque: revisited with functional magnetic resonance imaging. *J. Neurosci.* 21:8594–601
127. Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R. 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:187–203
128. Kourtzi Z, Kanwisher N. 2001. Representation of perceived object shape by the human lateral occipital complex. *Science* 293:1506–9
129. Desimone R, Schein SJ. 1987. Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *J. Neurophysiol.* 57:835–68
130. Felleman DJ, Van Essen DC. 1983. The connections of area V4 of macaque extrastriate cortex. *Soc. Neurosci. Abstr.* 9:153
131. Steele GE, Weller RE, Cusick CG. 1991. Cortical connections of the caudal subdivision of the dorsolateral area (V4) in monkeys. *J. Comp. Neurol.* 306:495–520
132. Maunsell JH, Van Essen DC. 1983. The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J. Neurosci.* 3:2563–86
133. Ungerleider LG, Desimone R. 1986. Cortical connections of visual area MT in the macaque. *J. Comp. Neurol.* 248:190–222
134. Polonsky A, Blake R, Braun J, Heeger DJ. 2000. Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nat. Neurosci.* 3:1153–59
135. Gandhi SP, Heeger DJ, Boynton GM. 1999. Spatial attention affects brain activity in human primary visual cortex. *Proc. Natl. Acad. Sci. USA* 96:3314–19
136. Tong F, Engel SA. 1910. Interocular rivalry revealed in the human cortical blindspot representation. *Nature* 411:195–99
137. Kastner S, Ungerleider LG. 2000. Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* 23:315–41
138. Blake R, Logothetis NK. 2002. Visual competition. *Nat. Rev. Neurosci.* 3:13–23
139. Luck SJ, Chelazzi L, Hillyard SA, Desimone R. 1997. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* 77:24–42
140. McAdams CJ, Maunsell JH. 1999. Effects of attention on orientation-tuning

functions of single neurons in macaque cortical area V4. *J. Neurosci.* 19:431–41

141. Kastner S, De Weerd P, Desimone R, Ungerleider LG. 1998. Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science* 282:108–11
142. Brewer AA, Press W, Logothetis NK, Wandell B. 2002. Visual areas in macaque cortex measured using functional MRI. *J. Neurosci.* 22:10416–26
143. Disbrow E, Roberts TP, Slutsky D, Krubitzer L. 1999. The use of fMRI for determining the topographic organization of cortical fields in human and nonhuman primates. *Brain Res.* 829:167–73
144. Disbrow EA, Slutsky DA, Roberts TP, Krubitzer LA. 2000. Functional MRI at 1.5 tesla: a comparison of the blood oxygenation level-dependent signal and electrophysiology. *Proc. Natl Acad. Sci. USA* 97:9718–23
145. Wandell B, Wade A. 2003. Functional neuroimaging of the visual pathway. *Neurol. Clin.* 21:417–43
146. Mata M, Fink DJ, Gainer H, Smith CB, Davidsen L, et al. 1980. Activity-dependent energy metabolism in rat posterior pituitary primarily reflects sodium pump activity. *J. Neurochem.* 34:213–15
147. Chiu SY, Kriegler S. 1994. Neurotransmitter-mediated signaling between axons and glial cells. *GLIA* 11:191–200
148. Mosier K, Bereznaya I. 2001. Parallel cortical networks for volitional control of swallowing in humans. *Exp. Brain Res.* 140:280–89
149. Tettamanti M, Paulesu E, Scifo P, Maravita A, Fazio F, et al. 2002. Interhemispheric transmission of visuomotor information in humans: fMRI evidence. *J. Neurophysiol.* 88:1051–58
150. Wolff HG. 1938. The cerebral blood vessels: anatomical principles. In *The Circulation of the Brain and Spinal Cord*, pp. 29–67. Baltimore: Williams & Wilkins
151. Reina-De L, Rodriguez-Baeza A, Sahuquillo-Barris J. 1998. Morphological characteristics and distribution pattern of the arterial vessels in human cerebral cortex: a scanning electron microscope study. *Anat. Rec.* 251:87–96
152. Chambers R, Zweifach BW. 1944. Topography and function of the mesenteric capillary circulation. *Am. J. Anat.* 75:173–205
153. Ravens JR. 1974. Anastomoses in the vascular bed of the human cerebrum. In *Pathology of Cerebral Microcirculation*, ed. J Cervos-Navarro, pp. 26–32. Berlin/New York: De Gruyter
154. Iadecola C, Yang G, Ebner TJ, Chen G. 1997. Local and propagated vascular responses evoked by focal synaptic activity in cerebellar cortex. *J. Neurophysiol.* 78:651–59
155. Woolsey TA, Rovainen CM, Cox SB, Henegar MH, Liang GE, et al. 1996. Neuronal units linked to microvascular modules in cerebral cortex: response elements for imaging the brain. *Cereb. Cortex* 6:647–60
156. Hyder F, Renken R, Kennan RP, Rothman DL. 2000. Quantitative multi-modal functional MRI with blood oxygenation level dependent exponential decays adjusted for flow attenuated inversion recovery (BOLDED AFFAIR). *Magn. Reson. Imaging* 18:227–35
157. Kida I, Kennan RP, Rothman DL, Behar KL, Hyder F. 2000. High-resolution CMR O_2 mapping in rat cortex: a multiparametric approach to calibration of BOLD image contrast at 7 Tesla. *J. Cereb. Blood Flow Metab.* 20:847–60
158. Heeger DJ, Huk AC, Geisler WS, Albrecht DG. 2000. Spikes versus BOLD: What does neuroimaging tell us about neuronal activity? *Nat. Neurosci.* 3:631–33
159. Heeger DJ, Ress D. 2002. What does fMRI tell us about neuronal activity? *Nat. Rev. Neurosci.* 3:142–51
160. Engel SA, Glover GH, Wandell BA. 1997. Retinotopic organization in human visual

cortex and the spatial precision of functional MRI. *Cereb. Cortex* 7:181–92
161. Liu HL, Pu Y, Nickerson LD, Liu Y, Fox PT, Gao JH. 2000. Comparison of the temporal response in perfusion BOLD-based event-related functional MRI. *Magn. Reson. Med.* 43:768–72
162. Boynton GM, Engel SA, Glover GH, Heeger DJ. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* 16:4207–21
163. Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, et al. 1994. fMRI of human visual cortex. *Nature* 369:525
164. Wandell BA, Press WA, Brewer A, Logothetis NK. 2000. *Soc. Neurosci. Abstr.* 26:821

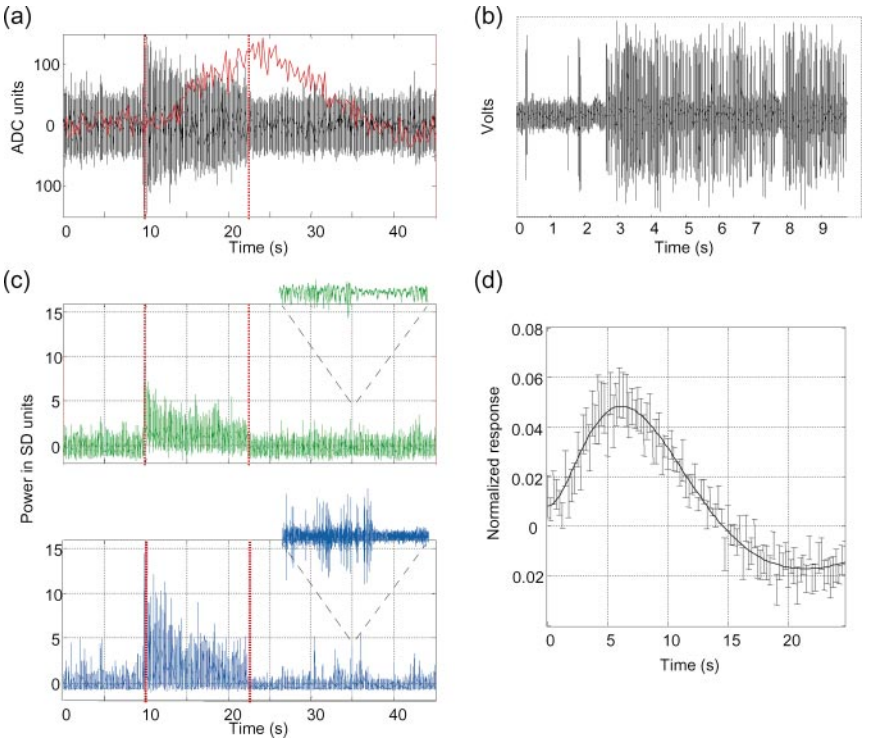

Figure 2 Simultaneous measurements of the neural and BOLD signals. (*a*) The black trace is a comprehensive (full-bandwidth; 0.050 Hz–22.3 kHz) mEFP signal. The red trace shows the average BOLD response measured by combining approximately 25 voxels ($1 \times 1 \times 2$ mm) near the electrode tip. The vertical red dotted lines delimit the stimulus presentation period. (*b*) Spike activity derived from the mEFP. (*c*) Frequency band separation of the mEFP. The green trace (*upper panel*) is the LFP. The signal is extracted by band pass filtering (10–150 Hz), rectification, and resampling (500 Hz). The blue trace (*lower panel*) is the multiunit activity (MUA). The signal is extracted from mEFP by band pass filtering (500–3000 Hz), rectification (absolute value), and resampling (1 Hz). The distinct characteristics of LFP and MUA signals before rectification are given in the expanded time scale of the insets, which show an interval of 1 s. (*d*) Estimated temporal impulse response function relating the neurophysiological and BOLD measurements in monkey (81). Very similar estimates are obtained using the LFP or MUA as the input neurophysiological signal.

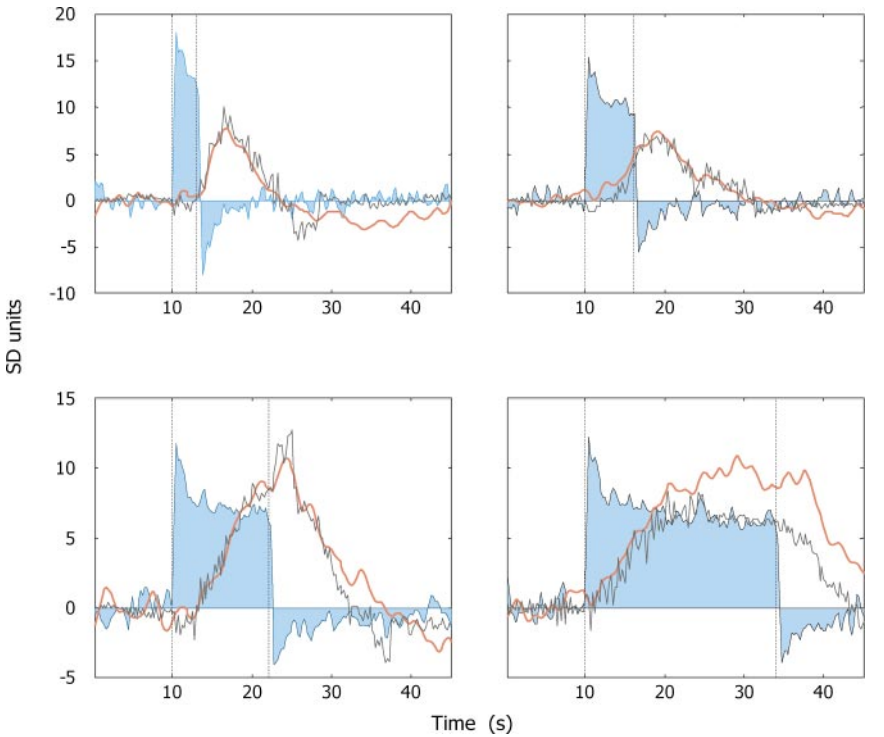

Figure 3 Simultaneous fMRI and intracortical recordings for visual stimuli of different duration. The four panels show responses to stimuli of 3, 6, 12, and 24 s. The blue-shaded trace is the LFP; the red trace is the BOLD response; the gray trace is the predicted BOLD response after convolution of the LFP with the temporal impulse response function. Linear time invariant (LTI) predictions are in good agreement over the duration range from 3 to 12 s, but the predictions of the latter portion of the 24-s stimulus are poor, suggesting the existence of nonlinearities not captured by the LTI analysis. The failure of accurate estimation of longer responses is in agreement with reports from human data, as described in the text.

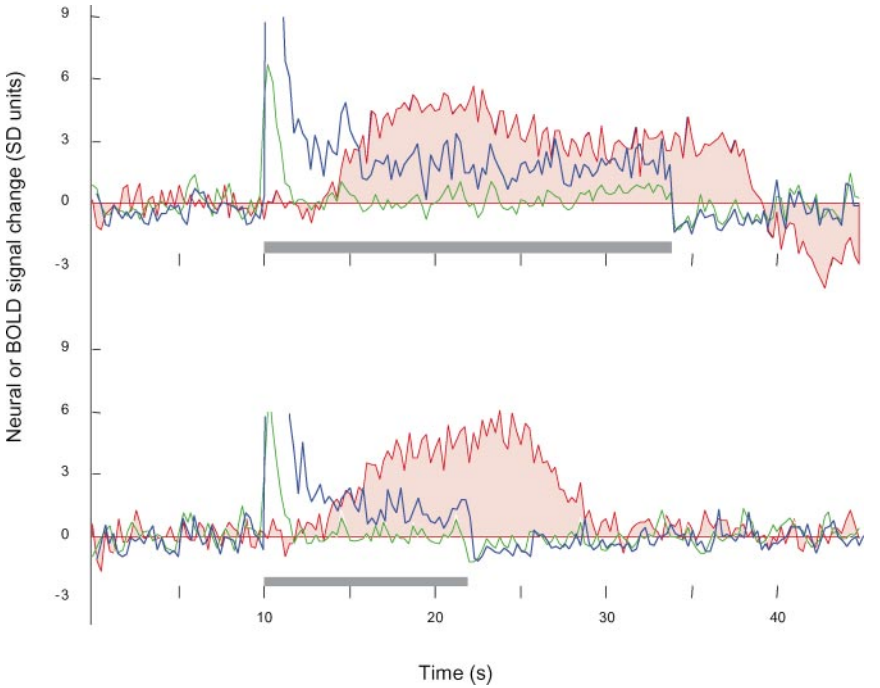

Figure 4 A cortical site in which the LFP and MUA predictions of the BOLD response differ significantly. Responses to a 24-s (*upper*) and 12-s (*lower*) stimulation are shown. The blue trace measures the LFP, the green trace measures the MUA, and the red-shaded trace measures the BOLD response. The MUA signal is brief and approximately the same in both experiments. The LFP and BOLD responses covary with the stimulus duration. At this cortical site and in 25% of the measurements, the LFP response matched the BOLD response but the MUA did not. There were no instances in which the MUA matched the BOLD response, but the LFP did not.

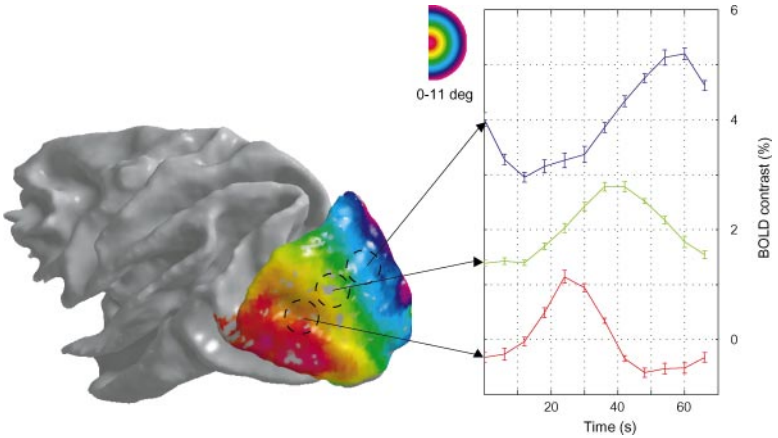

Figure 6 BOLD estimates of retinotopic organization in macaque V1 agree with single-unit estimates. The visual stimuli were a series of slowly expanding rings, each containing a collection of flickering squares (160, 163, 164). The stimulus begins as a small spot located at the center of the visual field; the spot becomes an expanding ring that grows to the edge of the stimulus display. As the ring disappears from view, a new spot, starting at the center replaces it. This stimulus causes a traveling wave of neural activity beginning in the foveal representation, several millimeters posterior to the lunate sulcus. Measuring the time course of the BOLD responses at a series of 3-mm radius regions-of-interest, we see that the BOLD response near the foveal representation is phase-advanced compared with the BOLD response measured in the more peripheral representations (*graph at right*). The phase of the BOLD response measures the visual field eccentricity that most effectively stimulates each cortical location. The most effective visual field eccentricity is indicated by pseudo-coloring the cortical surface according to the inset at the top (adapted from 142).

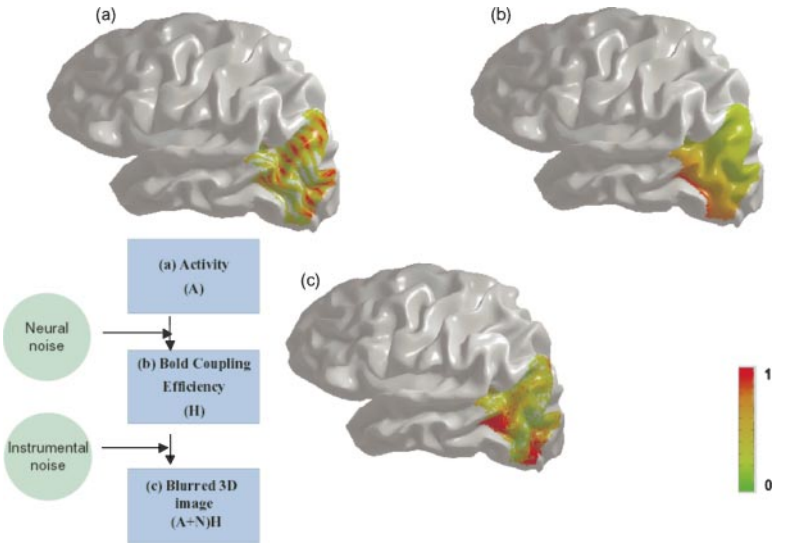

Figure 7 A computational model of the BOLD response. The inset in the lower left summarizes the sequence of steps used to create the three simulation images. Human cortex is segmented and approximated with a geometrically defined surface; the surface is presented as slightly inflated in order to expose the sulci. (*a*) Simulated neural activity, shown at its true spatial resolution. The neural activity varies at fine spatial resolution in the dorsal-ventral direction and at a coarser resolution in the posterior-anterior direction. (*b*) A representation of the coupling between the neural response and the BOLD response (HRE). In this simulation, the HRE is high on the lateral surface and low on the dorsal surface. (*c*) The product of the neural activity and HRE is blurred using a 5-mm (full-width half height) Gaussian kernel. The blurring is performed along the cortical surface (160). The pseudo-color overlay codes the relative size of the simulated measurements (arbitrary units; *key at right*).

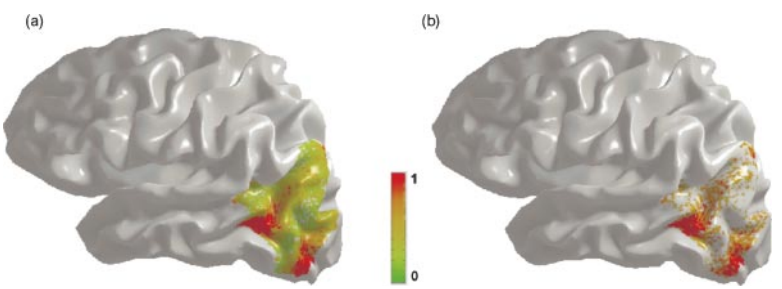

Figure 8 Simulations of the differential BOLD response. (*a*) The simulated differential BOLD response to the neural activity in Figure 6*a*. (*b*) The same BOLD response is shown after thresholding with a relatively high criterion. Other details as in Figure 7.

Annual Review of Physiology
Volume 66, 2004

CONTENTS