Review

Spatial and temporal limits in cognitive neuroimaging with fMRI

Ravi S. Menon and Seong-Gi Kim

A large body of research in human perception and cognition has been concerned with the segregation of mental events into their presumed hierarchical processing stages, the temporal aspect of such processing being termed 'mental chronometry'. Advances in single-event functional magnetic resonance imaging (fMRI) have allowed the extraction of relative timing information between the onset of activity in different neural substrates as well as the duration of cognitive processing during a task, offering new opportunities in the study of human perception and cognition. Single-event fMRI studies have also facilitated increased spatial resolution in fMRI, allowing studies of columnar organization in humans. Important processes such as object recognition, binocular vision and other processes are thought to be organized at the columnar level; thus, these advances in the spatial and temporal capabilities of fMRI allow a new generation of cognitive and basic neuroscience studies to be performed, investigating the temporal and spatial relationships between these cortical sub-units. Such experiments bear a closer resemblance to single-unit or evoked-potential studies than to classical static brain activation maps and might serve as a bridge between primate electrophysiology and human studies. These advances are initially demonstrated only in simple visual and motor system tasks and it is likely to be several years before the techniques we describe are robust enough for general use.

Mental events involved in sensory or cognitive processing take time. Even the seemingly simple act of reaching for a pen involves many tens of milliseconds from the time the image of the object is formed on the retina to our first conscious perception of the object, a similar delay as object attributes are processed and several hundred milliseconds more before motor movements commence. Over the past century, a large body of research in human perception and cognition has been concerned with the dissection of the brain activity during such tasks into presumed hierarchical processing stages, a concept known as mental chronometry[1,2]. Mental chronometric tasks have been used extensively in cognitive science[2] and cognitive neuroscience[3–5] to elucidate indirectly mechanisms underlying cognitive processing, often using reaction time (RT) as a measure of the processing load. As a very simple example, it has been known for over a hundred years[6] that the RT for detection of a flash of light depends on its intensity, but it is not known how or where this delay is distributed in the processing chain leading from the retina to the manual key press that indicates detection. Thus, while psychophysical measurements yield considerable information, they do not directly reveal the neural substrates or pathways involved in the processing of the stimulus.

Most brain imaging relies on the notion that distinct regions 'light up' (show increased activity) during particular tasks. That discrete substrates actually underlie information processing in the brain hinges on the second concept we wish to introduce, namely that the brain is segmented into many distinct areas that are specialized for their functional roles[7]. Numerous electrophysiological, autoradiographic and blood flow studies have shown that the brain does indeed have such spatially segmented areas that are functionally specialized. These are organized at different scales, from the scale of hemispheres (e.g. speech dominance), to that of gyri (e.g. primary sensory and motor cortex) down to cortical columns (e.g. ocular dominance columns). A considerable amount of early sensory processing, be it in the visual,

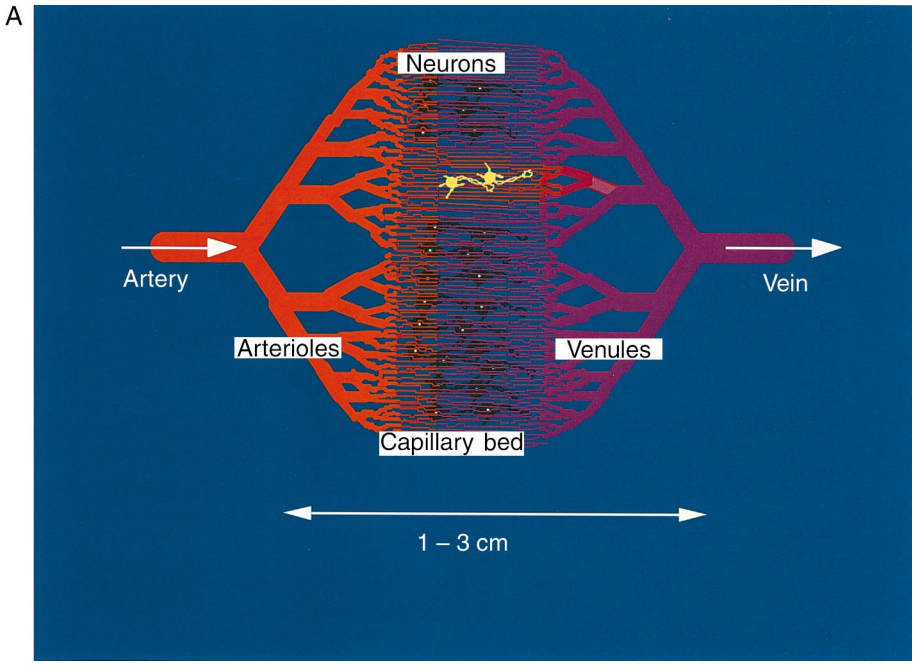
R.S. Menon is at the Laboratory for Functional Magnetic Resonance Research, The John P. Robarts Research Institute, PO Box 5015, 100 Perth Drive, London, Ontario, Canada N6A 5K8. S-G. Kim is at the Center for Magnetic Resonance Research, University of Minnesota Medical School, 2021 6th Street SE, Minneapolis, MN 55455, USA.

tel: +1 519 663 5777, **ext.** 4148
fax: +1 519 663 3403
e-mail: rmenon@irus.rri.on.ca

1364-6613/99/$ – see front matter PII: S1364-6613(99)01329-7

Box 1. What are we measuring?

When neurons fire in response to sensory or cognitive process, the glial cells, the nerve cell bodies and the synaptic terminals of the axon perform a carefully orchestrated sequence of events resulting in graded action potentials being transmitted and received (yellow 'active' cells in Fig. IA). It is thought that one consequence of this heightened activity is an increase in the local cerebral metabolism brought about primarily by the successive stages of delivery of the neurotransmitters to the synapse, clearing the neurotransmitter from the synaptic cleft, recycling it and repackaging it. Unfortunately, fMRI does not directly detect the electrical activity, nor does it measure the rapid increase in metabolism. Rather, it measures the increase in regional cerebral blood flow (rCBF) in response to the increased metabolism. Of course, tonic electrical activity exists in most regions of the brain simultaneously, and regional differences in this activity have been observed with PET and it is important to realize that increases and decreases in local brain activity are superimposed on this background activity. A recent review (Ref. a) explores the relationship between synaptic activity, rCBF and metabolic load in greater detail. Although the induced rCBF increase is an accepted marker for the functional electrical activity, the mechanism of activation-related rCBF control is not yet established.



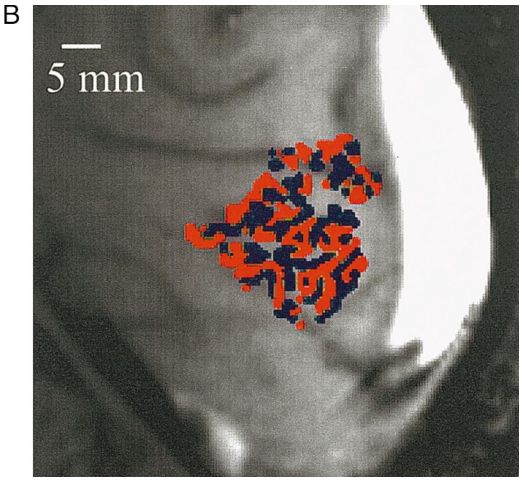


Fig. I. (A) Vasculature in the brain. The various sizes of cerebral vessels and the neuronal network within the capillary bed are shown. Activity of neurons (shown here in highlighted yellow neurons) is thought to result in an increase in the local cerebral blood flow, which is measured by fMRI. **(B)** Midline sagittal fMRI map and image of the right bank of the occipital pole of a normal volunteer showing ocular dominance stripes, corresponding to the left and right eye inputs (red and blue, respectively).

The process by which the MRI signal reflects increased brain activity is not completely characterized either, being related in subtle ways to the blood volume, blood flow, blood vessel geometry and oxygen consumption (Refs b–f). The fundamentals are well established, being based on a phenomenon known as the Blood Oxygenation Level Dependent (BOLD) effect (Refs e,g,h). In response to the activation related events described, the rCBF increases to the relevant region, but for reasons that are still not well understood, the rCBF increases far more than the expected increase in oxygen demand. This gives rise to the paradoxical situation in which the oxygenation state of the local capillary and venule beds is higher during focal brain activity than during rest. This was first observed visually by Penfield over 60 years ago, who noted that the venous blood became redder (that is, closer to the color of arterial blood) during localized seizures. More recently, Ogawa and colleagues, relying on the paramagnetic properties of the deoxyhemoglobin present in the capillaries and veins, demonstrated that MRI could be made sensitive to the local oxygenation state of the blood as well (Refs g,h). As the blood becomes redder,

auditory or motor cortex, is accomplished by columnar subunits that interact in complex ways. The parcellation of the visual cortex into multiple specialized areas (layers, columns and areas such as V1–V8) represent perhaps the best examples of the use of fMRI to study functional segregation[8–11], and evolve primarily from individuals who previously pursued such studies using deoxyglucose, cytochrome oxidase and electrophysiological techniques in non-human primates.

In this review we will explore the use of fMRI in elucidating the role of brain areas involved in processing certain types of information on a combined spatial-temporal scale never before imagined. The advances we discuss are made possible by developments in single-event fMRI applied at very high magnetic field strengths. While the concept of acquiring averaged and time-locked fMRI acquisitions to brief stimulus presentations is not new in sensory studies[12–14], single-event cognitive studies are more recent[15–17]. What constitutes a single event? In the present context, it refers to a single cognitive, sensory or motor event such as a single finger movement, a single presentation of a pair of objects in a forced choice experiment or the presentation of a brief visual stimulus. The single-event fMRI data might then be averaged, often by using a behavioral correlate (such as movement onset) to align the fMRI response. Temporal accuracy with averaged trials can be as stable as ±50 ms in

the MRI signal in appropriately acquired images increases. Because the arterial side of the brain vasculature is essentially fully oxygenated, no BOLD changes occur on this side. The BOLD effect is manifested in the bluish capillary beds, venules and draining veins, which are only 60–70% saturated with oxygen at rest and hence have the capacity to get 'redder', with the corresponding increase in MRI signal intensity.

Several major issues limit reaching arbitrarily high spatial resolution in fMRI, not the least of which is a lack of signal. As shown in Fig. IA, the size of the vasculature giving rise to the observed MRI signal changes can vary from the capillary bed in the cortex (<10 μm) to draining veins (a few mm) (Refs i,j,k). Signal changes in the capillary bed are more accurately co-localized with the neural site of activity, but those in the draining veins can be centimeters away and hence are not desirable (Ref. l). Thus the microvascular sensitivity that is associated with very high field scanners will probably be necessary to map at the columnar level. As a corollary, higher spatial resolution appears to require higher intrinsic signal, giving a potential advantage to the higher field research scanners (Ref. m). An example of current capabilities at 4 Tesla (T) is shown in Fig. IB (Ref. n). The ocular dominance stripes, corresponding to the left and right eye inputs, shown in this image, look very much like those seen in optical imaging (Ref. o) or post mortem studies (Ref. p). Such maps can only be acquired if the MRI technique is sufficiently sensitive to the microvasculature, if the microvascular response is confined to active neurons (yellow cells) and the macrovascular signal can be controlled for. Controlling the macrovascular response is achieved by the brief presentation of stimuli rather than the use of block designs and the early part of the hyperoxygenation phase (Ref. n). Optical imaging work has shown that short duration stimuli lead to more consistent responses with fewer vascular artifact than long duration stimuli (Ref. q). Thus, the use of single-event paradigms has facilitated very high resolution cortical mapping as well.

Because inhibition and excitation are both energy consuming processes, it is not clear that they can be differentiated using the fMRI response in any single region. For example, one probably needs to look distal to the site of inhibition to see a decrease in the fMRI signal intensity. This implies evaluating the brain as a network, or at the very least, knowing something about the relevant connectivity (Ref. r). Similarly, we do not know what relative contribution spiking and subthreshold action potentials make to the fMRI measurements of a region. Ultimately this may also be important for spatial localization, because subthreshold activity often extends further in space than the active spiking (Ref. o).

Finally, the concept of mental chronometry needs to be dealt with in the context of connectivity between nodes in this neural network. Chronometry implies that neural substrates activate in a temporal order. Clearly this is only valid in a hierarchical model of information processing, and while there is evidence that some neural networks behave this way, there is also considerable evidence that others do not. The brain is known to process information in parallel, to exchange information via feedforward and feedback loops and to convey information by synchronized oscillatory coupling. The latter is particularly important, but would produce no phase lag between the two communicating nodes. A possibility for detecting such oscillating effects might include very rapid imaging with complex space-time analyses, or one might be lucky enough to find that the BOLD signal amplitude depends on the coherence of the oscillating neural assembly (Ref. s).

References

a Jueptner, M. and Weiller, C. (1995) Does measurement of regional cerebral blood flow reflect synaptic activity? – implication for PET and fMRI *NeuroImage* 2, 148–156

b Woolsey, T.A. *et al.* (1996) *Dynamic Measurements of Local Cerebral Blood Flow*, Academic Press

c Ogawa, S. *et al.* (1993) Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging: A comparison of signal characteristics with a biophysical model *Biophys. J.* 64, 803–812

d Thulborn, K.R. (1998) A BOLD move for fMRI *Nat. Med.* 4, 155–156

e Ogawa, S. *et al.* (1998) On the characteristics of fMRI in the brain *Annu. Rev. Biophys. Biomol. Struct.* 27, 447–474

f Davis, T.L. *et al.* (1998) Calibrated functional MRI: mapping the dynamics of oxidative metabolism *Proc. Natl. Acad. Sci. U. S. A.* 95, 1834–1839

g Ogawa, S. *et al.* (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation *Proc. Natl. Acad. Sci. U. S. A.* 87, 9868–9872

h Ogawa, S. *et al.* (1990) Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high fields *Magn. Reson. Med.* 14, 68–78

i Menon, R.S. *et al.* (1993) 4 Tesla gradient recalled echo characteristics of photic stimulation induced signal changes in the human primary visual cortex *Magn. Reson. Med.* 30, 380–386

j Lai, S. *et al.* (1993) Identification of vascular structures as a major source of signal contrast in high resolution 2D and 3D functional activation imaging of the motor cortex at 1.5 T: Preliminary results *Magn. Reson. Med.* 30, 387–392

k Lee, A.T., Glover, G.H. and Meyer, C.H. (1995) Distribution of large venous vessels in time-course spiral blood-oxygen-level-dependent magnetic resonance functional neuroimaging *Magn. Reson. Med.* 33, 745–754

l Kim, S.G. *et al.* (1994) Potential pitfalls of functional MRI using conventional gradient-recalled echo techniques *NMR Biomed.* 7, 69–74

m Gati, J.S. *et al.* (1997) Experimental determination of the BOLD field strength dependence in vessels and tissue *Magn. Reson. Med.* 38, 296–302

n Menon, R.S. and Goodyear, B.G. (1999) Submillimeter functional localization in human striate cortex using BOLD contrast at 4 Tesla: implications for the vascular point spread function *Magn. Reson. Med.* 41, 230–235

o Das, A. and Gilbert, C.D. (1995) Long-range horizontal connections and their role in cortical reorganization revealed by optical recording of cat visual cortex *Science* 375, 780–784

p Horton, J.C. *et al.* (1991) Arrangement of ocular dominance columns in human visual cortex *Arch. Ophthalmol.* 108, 1025–1031

q Malonek, D. and Grinvald, A. (1996) Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy implications for functional brain mapping *Science* 272, 551–554

r Friston, K.J. (1997) Imaging cognitive anatomy *Trends Cognit. Sci.* 1, 21–27

s Singer, W. (1995) Development and plasticity of cortical processing architectures *Science* 270, 758–764

certain cases[18], while true single trials can show responses separated by a second or two[19,20]. We will examine the limitations of both the spatial and temporal resolution of fMRI technique and show how, in certain situations, these can be dramatically extended.

Cinematography using fMRI

Because cerebral hemodynamic responses to brain activity are quite sluggish, brain mapping techniques based on hemodynamic properties have never been seriously considered as candidates for chronometric examination of the human brain. Although regional cerebral blood flow (rCBF) was hypothesized to be proportional to neuronal activity over a century ago[23], it was not until the advent of positron emission tomography (PET), that rCBF changes in response to modulation of neural activity could be efficiently mapped with good spatial resolution in humans using radiolabelled water[7,24]. PET revolutionized the approach to systems-level neuroscience and cognition by allowing one to examine the brain foci involved in a given cognitive task. A significant limitation of PET (and fMRI as generally practiced) is that the maps produced are static representations of the dynamic activity of the brain averaged over a long time period relative to the mental processing scale[25]. This is because the majority of neuroimaging studies to date have used so-called block designs, in which the behavior is performed repeatedly

Box 2. Temporal features of the fMRI response

Figure I shows the time course of the fMRI signal change in primary visual cortex (V1) in response to a typical visual stimulus of 10 s duration (shown as box from 2 to 12 s), measured as the average of 54 single trials at a magnetic field strength of 1.5 Tesla (T). Several time points (i–iv) can be used to characterize the temporal aspects of the fMRI measured hemodynamic response. Firstly, it is noted that the fMRI signal does not start to change until just over 2 s after the onset of the stimulus [delay (i)], while the time-to-peak of the response [delay (ii)] is about 13 s. Furthermore, the total vascular response takes a considerable amount of time to return to zero [delay (iv)] and stabilize after the stimulus is turned off. The duration of this recovery places limits on how fast the stimulus can be repeated. For a stimulus duration of up to about 4 s, the time-to-peak depends on the duration and other parameters of the stimulus. Beyond 4 s, the neural (and hence BOLD) responses saturate, and further increases in stimulus duration are manifested solely in the width of the peak, with some as yet not understood variations. It has been determined that the inflection point from baseline can be used as a measure of relative stimulus onset (Refs a–d), while the width of the response [delay (iii)] can be used as a measure of stimulus processing time (Ref. e).

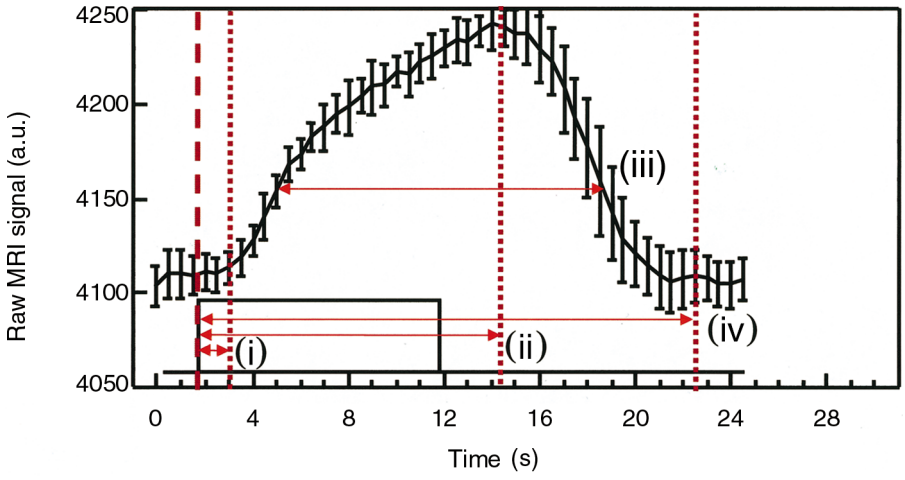


Fig. I. The time course of fMRI signal change in visual cortex in response to a visual stimulus. (See text.) (Figure kindly provided by Dr Robert Savoy.)

Savoy observed that the variability of the onset latencies in simple visual experiments was small (Refs a,e) and visual inspection of his averaged data kindly provided for this figure (with standard deviation bars derived from the 54 averaged trials) suggested that the rising edge would be the most precise indicator of onset, in terms of smallest onset-latency variability. The falling edge of the response seemed particularly variable from trial to trial. In our own work, we have also observed these phenomena using different graphical techniques (Ref. d). It should be obvious that limiting the analysis to just the initial rapid rise in fMRI-detected response throws away a good deal of analytical power, since one rarely has more than a few dozen points along the rising edge. In a context where that additional power is needed (always!), it makes much more sense to use the entire response function, although some assumptions about the underlying neural and hemodynamic responses need to be made. Many groups are pursuing such modeling for a variety of different reasons (Refs f,g).

As pointed out in Box 1, fMRI is sensitive to vasculature ranging in size from capillaries to venous sinuses and as one might imagine, the draining vasculature distorts the measurement of timing. MRI experiments performed at high temporal resolution allow the potential separation of the microvascular and macrovascular responses based on timing, because the veins 'activate' several seconds later than the parenchyma as the bolus of oxygenation-altered blood originating in the capillaries flows through the venous system.

References

a Savoy, R.L. *et al.* (1995) Pushing the temporal resolution of fMRI: Studies of very brief visual stimuli, onset variability and asynchrony, and stimulus-correlated changes in noise *Soc. Magn. Reson. Abstr.* 1, 450

b Kim, S-G., Richter, W. and Ugurbil, K. (1997) Limitations of temporal resolution in functional MRI *Magn. Reson. Med.* 37, 631–636

c Richter, W. *et al.* (1997) Time-resolved fMRI of mental rotation *NeuroReport* 8, 3697–3702

d Menon, R.S., Luknowsky, D.L. and Gati, J.S. (1998) Mental chronometry using latency-resolved functional magnetic resonance imaging *Proc. Natl. Acad. Sci. U. S. A.* 95, 10902–10907

e Savoy, R.L. *et al.* (1994) Exploring the temporal boundaries of fMRI: Measuring responses to very brief visual stimuli *Soc. Neurosci. Abstr.* 20, 1264

f Josephs, O., Turner, R. and Friston, K. (1997) Event-related fMRI *Hum. Brain Mapp.* 5, 1–6

g Rosen, B.R., Buckner, R.L. and Dale, A.M. (1998) Event-related functional MRI: Past, present and future *Proc. Natl. Acad. Sci. U. S. A.* 95, 773–780

(every few seconds) in a 'block' of trials and this is compared using a variety of statistical methods with another block of trials, typically (but not necessarily), a resting condition. Often the two types of blocks are interlaced, which allows powerful cross-correlation or statistical-parametric mapping techniques to be used in the analysis[25,26]. A simple example, used in both PET and fMRI, might be 30 s of finger tapping interleaved with 30 s of no movement with the pair repeated five times. This sort of study does not give any dynamic information.

In the early part of this 'Decade of the Brain', there was a dramatic and unanticipated extension of magnetic resonance imaging (MRI) to cognitive neuroimaging with the discovery that MRI images could also non-invasively map brain activity (Box 1). This MRI technique eliminated the use of exogenous contrast agents, ionizing radiation or radiotracers[27–29] and acquired the moniker of fMRI. Despite the ability to make MRI images on a modern MRI scanner in excess of 20 per second, the temporal dimension of fMRI has not been exploited for chronometric studies; in part because the complicated dependence of the fMRI signal on the coupling mechanism between neural activity and hemodynamic response is not well understood (Box 1), and perhaps more so, because of a feeling that rCBF, blood volume and oxygenation changes are too slow to be of much use in sub-second temporal studies of brain activity. Nonetheless, recent studies have demonstrated that while the rCBF change observed in fMRI is delayed by several seconds relative to the stimulus onset, both the onset of the fMRI response and the width of the fMRI response can be used to extract useful temporal information (Box 2).

If fMRI responses in all regions are identical, it is straightforward to compare temporal characteristics between regions. Using this assumption, researchers have attempted to

Box 3. Time-resolved event-related fMRI

To explain how vascular and neural delays can be separated, a hypothetical experiment is illustrated here (Fig. I). The two trials indicated here can be two true single trials (as in Fig. 1 and 2 of main article) or the average of many repeated trials where no time-dependent modulation is expected (as in Fig. 3 of main article). (**A**) We assume a cognitive task with three different neural components; one (shown as green) corresponds to the visual presentation of the task, the second (shown as red) corresponds to the cognitive process invoked by the task, and the last (shown as blue) corresponds to the motor component involved in responding to the task (such as a push button response or a joystick movement). The first visual component and the last motor component remain constant during repeated trials in our construct. But the cognitive processing component is made to vary in duration from trial to trial by externally imposing delay times or is the result of intrinsically different response times (RT). For simplicity, we assume the onset of cognitive processing is always at the same time. (**B**) The corresponding fMRI responses in the areas related to the first visual component are consistent during repeated trials. The area related to the motor component also has similar time courses since the motion duration for a button push is invariably the same, but its onset occurs at variable times after the stimulus presentation depending on processing load. The second component will induce fMRI responses with the same onset time, but different duration, depending on the task difficulty. (**C**) By correlating behavior (such as external delay or reaction time) with temporal characteristics of fMRI signals (such as onset time, dotted lines; and width, solid lines), the specific regions related to each of three functions can be mapped by regressing the behavioral correlate against the pixel intensity dependence from trial to trial. In this manner, pixels that exhibit similar temporal characteristics to the behavioral correlate can be found, and presumably, reflect activity that is involved in the processing strategy.

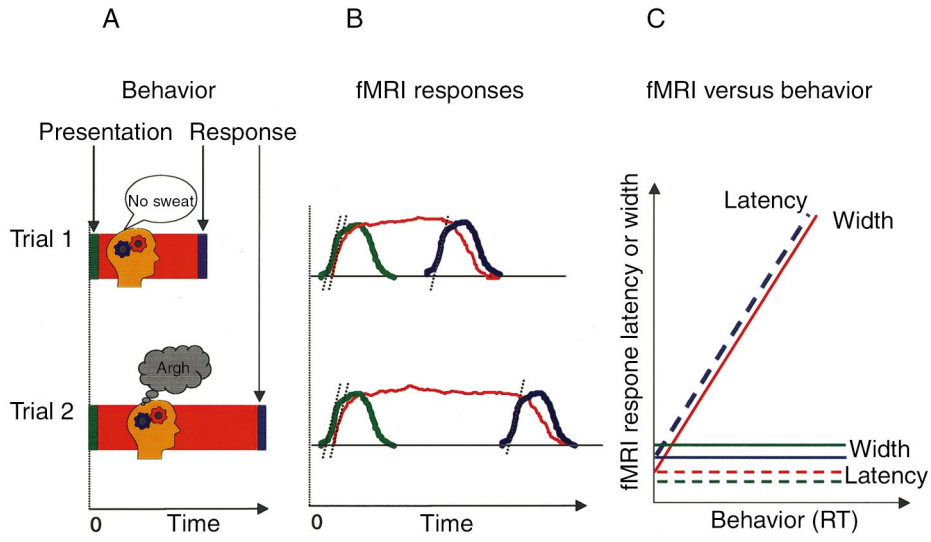


Fig. I. Time-resolved event-related fMRI: a hypothetical experiment. (See text.)

determine sequential neural processing using averaged event-related fMRI[15]. Buckner and colleagues acquired images gated to the onset of a task in a paradigm so that the temporal evolution of the fMRI signal during the execution of the task could be averaged following repeated executions of the same task. As they point out, if the intrinsic hemodynamic response among brain regions is different, as is undoubtedly the case, it is not possible to be conclusive about the temporal sequence of neural activity between different regions from the averaged time courses. For example, Buckner and colleagues observed that activation in the left prefrontal cortex language area was delayed about 2 s relative to activity in extrastriate areas during the performance of a word generation task. Their discussion suggests that a latency difference of that magnitude is likely primarily related to that of intrinsic hemodynamic responses of different vascular beds and not necessarily due to the order of activation of the different neural substrates in this cognitive task.

In general, if behavior and brain function do not vary during repeated trials, an averaging approach as used in the studies mentioned above is valid, because alignment of the fMRI response can be made to the presentation of the stimulus cue. However, if time-dependent modulation occurs (such as learning, alterations in strategy and errors, and habituation, all of which is common in cognition), averaging loses unique information associated with each *individual* execution of the task. Thus there is a category of tasks that can benefit from repeated averaging to build up sufficient signal-to-noise, and there are other tasks in which the effects of learning and strategy may lead to confounds in experimental design and interpretation. Careful behavioral measures must be performed to exclude the effect of the time-dependent modulation effects just mentioned.

The hemodynamic response is dynamic after all

To separate intrinsic hemodynamic differences from neural activity differences, a time-resolved event-related fMRI technique can be used (see Box 3). The idea is to examine how fMRI parameters vary with behavioral correlates and thus require multiple behavioral outcome measures (e.g. two different RTs). This might seem onerous, but at very high magnetic fields, there is often enough sensitivity to monitor fMRI signal evolution in a *single* execution of a cognitive task *without* averaging over many trials (see Figs 1 and 2). With this capability, it is possible to perform many such single-trial executions of a given task by collecting the data separately. If sensitivity is not high enough to detect significant activation during performance of a single trial, or if extreme temporal precision is desired, signal averaging can be performed using a behavioral correlate such as task performance (e.g. correct or incorrect) or response criteria (e.g. RT) to align the fMRI data prior to averaging. Subsequently, temporal characteristics (such as onset time and width shown in Box 2) of the fMRI responses can be correlated with behavioral data such as response width (Figs 1 and 2) or response onset time (Figs 2 and 3). In this way, differences in the basic temporal onset of neuronal activity

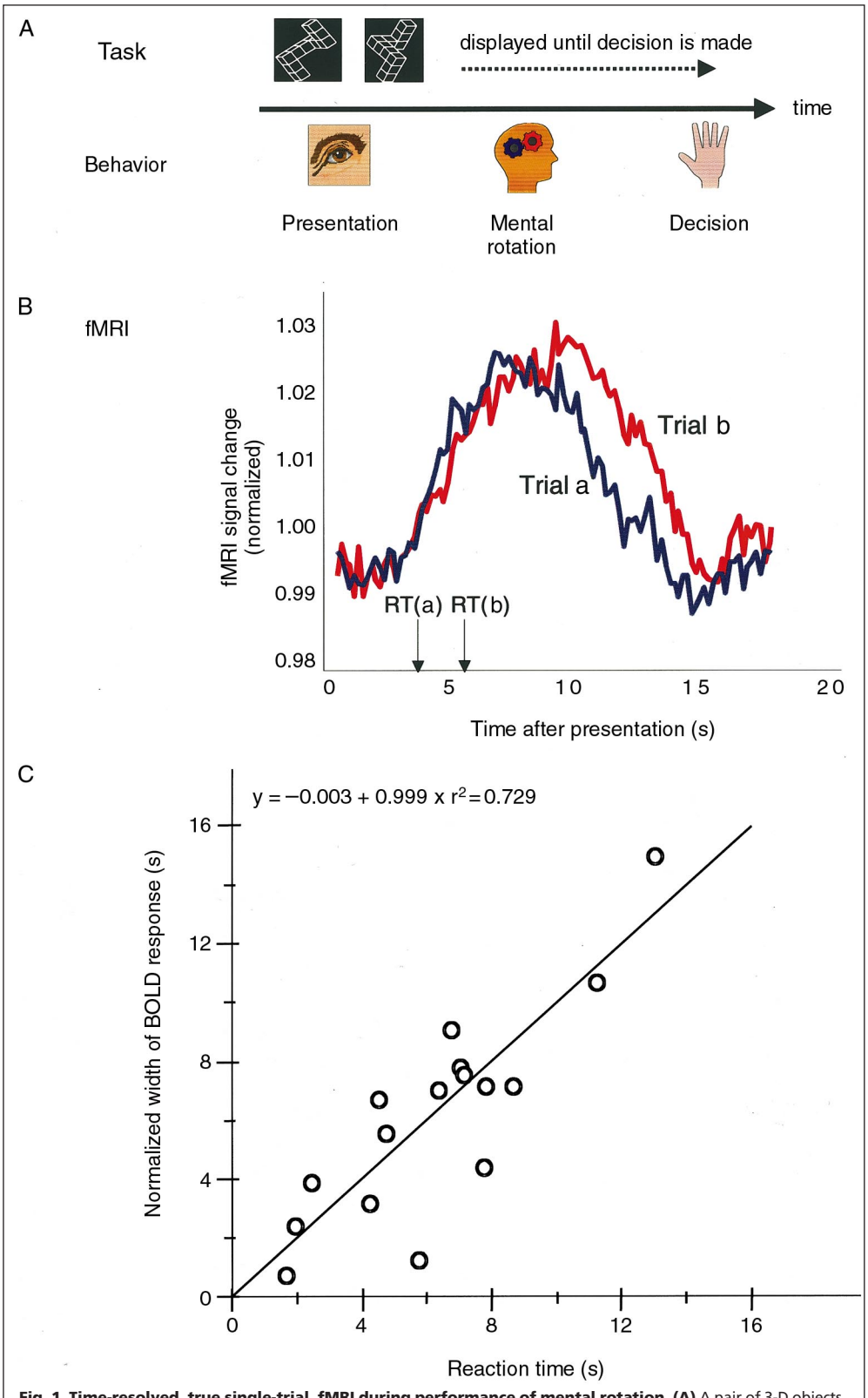


Fig. 1. Time-resolved, true single-trial, fMRI during performance of mental rotation. (A) A pair of 3-D objects similar to those used in the classic experiment by Shepard and Metzler[30] were shown to the subjects in the magnet and the subjects were asked to indicate via a button press if the two objects were identical (in general, at a different perspective or rotation angle), or mirror images of each other. For each subject, 16 single trials were performed with different pairs of objects. Reaction times (RT) varied from trial to trial and were recorded. Using simple t-tests, the superior parietal lobule (SPL) was found to be activated during performance of the mental rotation task, as were many other areas. To determine whether processing in the SPL was a constant or a variable of the task, the true single-trial time courses (without any averaging) of two trials (a and b) with the corresponding reaction times [RT(a), RT(b)] are illustrated in **(B)**. Although onset times of both trials are the same (presumably correlating with the start of mental activity), the width of fMRI response follows the reaction time. **(C)** Onset times and normalized durations (sum of rise time and plateau times) in all 16 trials were determined in one subject. The normalized width correlated well with RT, suggesting that the SPL is involved in mentally rotating the objects into registration.

(but not the structure of that neural activity) can be distinguished from hemodynamic response time variations between subjects and between brain areas. This approach allows the experimenter to obtain higher temporal resolution and so improve the determination of the function of the specific area compared to those data obtained from single averaged time courses which have been averaged without regard to behavioral correlates.

Mental rotation

Our first example is that of the well known comparison of two rotated objects. Shepard and Metzler[30] demonstrated that the time to decide whether two block-like objects were the same or different depended linearly on the rotation angle of one object relative to the other. This mirrors the situation found in Box 3, where the task varies from trial to trial (because of different rotation angles) and hence the RT also varies proportionally. One might then expect a component of the task whose width scales with duration of the mental rotation and another component of the task whose onset varies with RT. This is shown in Fig. 1. It was found that the width of the fMRI response in the superior parietal area was well correlated with the reaction time, and the onset time of the fMRI response in that region remained constant. This suggests that the superior parietal lobule is intimately involved in the mental rotation process[20,30], not just as a constant of the task, but as a substrate involved in the mental rotation. Conversely, it would be expected that the onset of activity in the motor area correlated with the RT for a button press indicating the objects were the same or different, because the button press involves a well practiced stereotyped movement whose duration probably does not vary significantly from trial to trial. In many cognitive tasks including mental rotation, neural processing lasts from a few hundred milliseconds to a few seconds and in this domain of applicability, fMRI may be used to examine sequential neural substrates in different regions.

Visuo-motor tasks

Another example of this procedure, where the processing duration is expected to vary, is given by fMRI responses of motor areas when compared with well-controlled motor preparation times[16,19]. Premotor (PM) and supplementary motor areas (SMA) were activated during motor preparation and execution periods (Fig. 2). Interestingly, primary motor cortex was also activated during both periods in this subject[16]. These observations are consistent with previous single-neuron recording studies in primates[3,31], suggesting that fMRI responses, albeit slow and blurred, follow neuronal activity with some fidelity. By using time-resolved true single-trial fMRI, the specific brain areas responsible for processing different components of a task can be determined during performance of cognitive operations.

Not all tasks have sufficiently robust activity to allow mapping in true single-trial fashion. In our final example, we demonstrate the use of onset time as an indicator of where the load associated with a motor tracking task may reside. In this task, 10 trials were averaged together, the onset of each trial being synchronized to the presentation of the task screen (Fig. 3). In this simple video game, akin to a similar set of experiments on monkeys[3,32], the RT delay could originate in the planning of the movement or in the execution of the movement. We note that the spread in RT seen in the results arises from differences in task performance between subjects and that there is no behaviorally measured time dependent modulation of function in this very well practiced task. Surprisingly, the data demonstrate that in similar age subjects, the hemodynamic responses scale consistently between subjects, and so RT *between* subjects can be used as a correlate in certain cases. In this case, it is shown that the inter-subject variation in RT arises from a delay somewhere between V1 and the SMA and not between the SMA and M1.

Although the above examples have been chosen to illustrate the limits of what is currently obtainable in terms of sensitivity and temporal resolution, one can reasonably expect that the number of institutions with the finances and expertise to support dedicated very high field scanners for cognitive research will be small. We anticipate that similar developments at a more accessible 1.5 T will still allow temporal characterization of the behavior via the fMRI signal. A number of clever methods in this regard have evolved very recently and underscore the importance of imaging scientists and neuroscientists working together to circumvent current limitations based on the history of their own fields. For example, Toni and coworkers have shown that by systematically varying the phase between behavioral events and image acquisition, one can

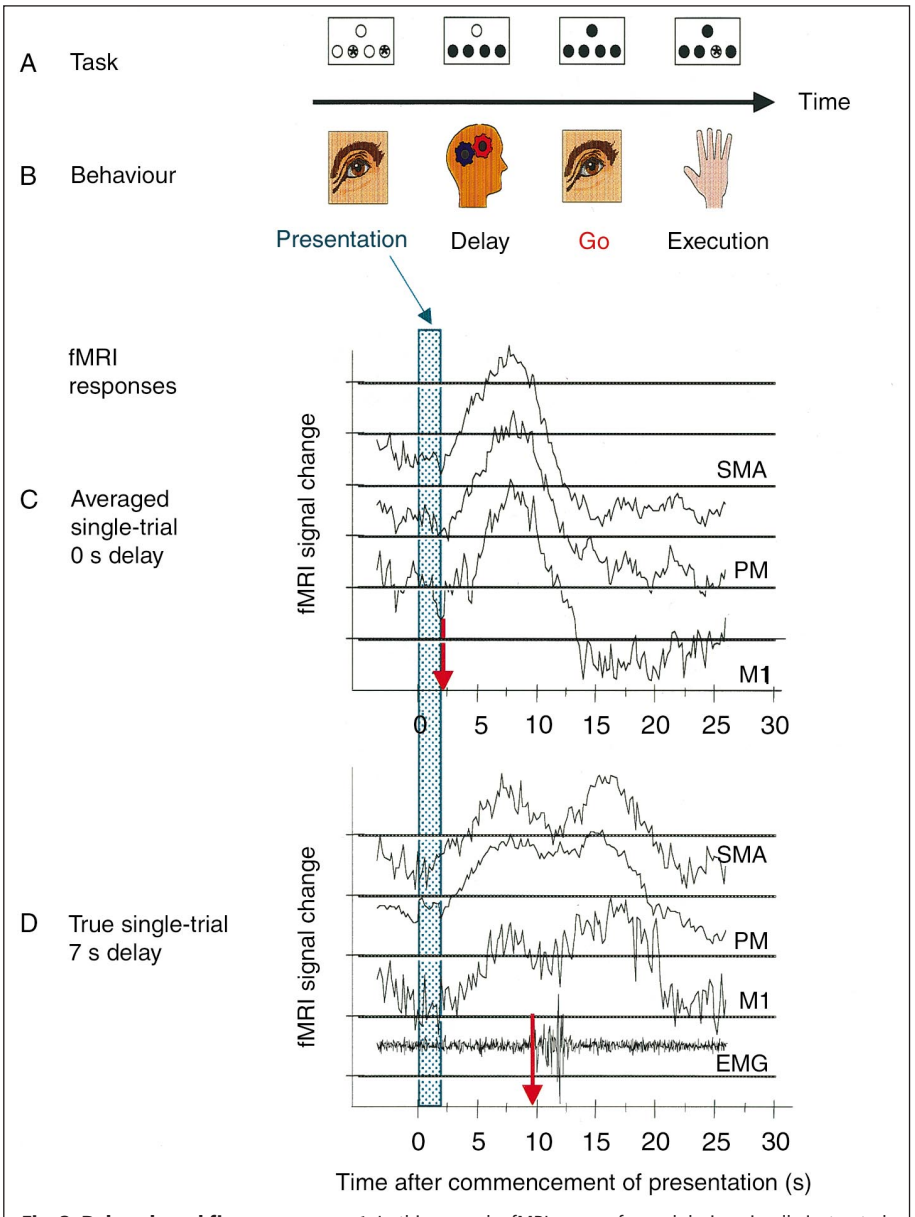


Fig. 2. Delayed cued finger movement. In this example, fMRI was performed during visually-instructed, cued four-finger movement task with an external delay time between visual instruction and cue. **(A)** The subject's view of the projected image. A row of four circles, one for each finger, was used for an instruction, and a fifth circle at the top of the screen was used as a commencement instruction (Go). **(B)** Each experiment consisted of three different periods. The first period was the presentation period [shown stippled blue in (C) and (D)]; the circles, initially empty, were sequentially filled, one every 700 ms, with a total presentation time of 2.1 s. The order of the filled circles defined the corresponding order of finger movements. The second period was a delay time, which was varied between 0 and 7 s. During this period, the four circles remained filled while the subject was (presumably) mentally preparing the finger movement. The last period was the execution period; after the Go circle [indicated by red arrows in (C) and (D)] was lit, the subject moved fingers in the order memorized. **(C)** Significant signal change of the averaged single-trial fMRI was observed in contralateral primary motor area (M1), bilateral premotor (PM), and bilateral supplementary motor area (SMA) during finger movement with the delay time of 0 s. **(D)** True single-trial, single-subject time courses in the three motor areas are shown for the experiment with 7 s delay. The EMG recording [bottom trace in (D)] shows absence of motion during the preparation period. The scale of the y-axis indicates fractional change relative to baseline for each individual time course shown. By comparing time courses with two delay times, all three motor areas involved in both the motor preparation and motor execution and their relative ordering can be discerned. Each division of the vertical scale represents 1% changes, but the fMRI responses have been offset relative to each other for clarity. The data demonstrate a two-component response in SMA, PM and M1. Immediately after the initial presentation of the sequence, activity rises in these areas, despite the fact that the EMG shows no muscle involvement, presumably indicating a motor set state. The activity slowly begins to die away during the delay period until the 'Go' signal is given, when an enhancement of the fMRI activity if again found as the motor pattern is executed, as evidenced by the EMG and button presses.

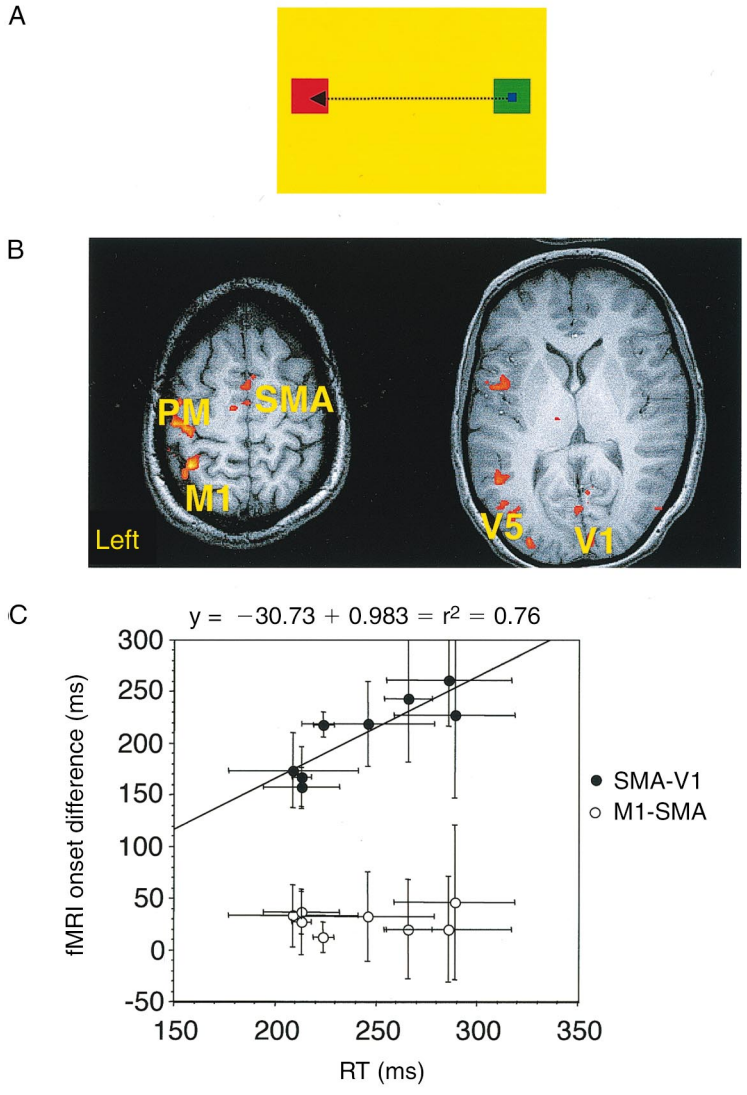


Fig. 3. Visuomotor reaction time experiment. This experiment demonstrates the use of onset time in determining where certain aspects of a simple stereotyped behavior originate[18]. **(A)** The visual stimulus that cued subject motor response is shown along with the direction of cursor movement. This stimulus appeared for 2 s every 30 s and the subject was instructed to use the joystick in their hand to move the blue cursor from the right side start box (green) to the left side target box (red) as rapidly and accurately as possible. The entire movement was digitized by a computer and each trial was repeated 10 times. **(B)** Activation maps in two slices generated using the cross-correlation approach of the areas invoked by the visuo-motor task in a single subject. The activation maps, are superimposed onto the corresponding anatomical slices. Abbreviations: PM, premotor area; M1, primary motor area; SMA, supplementary motor area; V5, the motion sensitive area of visual cortex; V1, primary visual cortex. These areas have been determined functionally and anatomically. **(C)** Plot of fMRI onset delay between V1 and SMA as a function of measured RT from the kinematic recording as well as the onset differences between SMA and M1 versus RT. We found that the delay between V1 activation and SMA activation scaled proportionally with RT across subjects, while the delay between SMA and M1 did not. Thus once the motor program is invoked, execution appears to be automatic. The delay in response appears to occur between V1 and the SMA, a finding that is consistent with electrophysiological recordings in monkeys[31,32].

still achieve high temporal resolution while scanning the whole brain[21]. This form of 'pseudo-temporal' resolution will undoubtedly be useful regardless of magnetic field strength, because we are already close to the physiological limits (auditory noise and magneto-stimulation) that the human body can tolerate in terms of image acquisition speed. Tasks such as working memory and mental rotation can also be done with conventional scanners, since these cognitive processes are often quite slow and do not require imaging to be pushed to its limits[33]. This is true for spatial-temporal resolution as well. Using a special MR scanner, Lee and colleagues have shown that not only can one dissociate preparation of finger movements from movement execution (similar to Fig. 2), but one can separate pre-SMA and SMA proper in terms of the spatial-temporal pattern of the fMRI response[22]. Even for tasks in which there is short-term adaptation, it is possible with sufficient averaging on multiple subjects to discriminate between neural substrates activated in the first 15 s of a movement and the last 15 s of that movement and to examine areas whose activity decreases with repetitive stereotyped movements[34]. Such schemes rely on a similar degree of initial naivete and adaptation strategy for subjects being averaged together, but if good behavioral measures are acquired in support of these assumptions, this should not pose a significant limitation.

How far can we go?

The developments discussed in this review open up the possibility of studying the temporal processing in the brain at the second and sub-second time scale, and the spatial processing of information at the sub-millimeter scale. We have concentrated on the relative onset of activity between different areas and the relative duration of activity deduced by the fMRI time courses. These are not the only aspects of timing in the brain, but they do represent the most challenging applications to which the technology has been pushed so far. Hundreds of papers presenting block designs have appeared over the past eight years since fMRI burst on the scene, and it is fair to state that modern clinical MRI scanners have achieved the stability to do these demanding serial studies that their predecessors did not. Several dozen papers using widely spaced single trials with averaging have also appeared in the past three years, and the importance of that methodology has significantly impacted paradigm design. Block designs, while currently out of fashion, will remain a staple of fMRI labs, regardless of magnetic field strength. As novel methods for deconvolving the hemodynamic response in closely spaced single trials are validated, one can anticipate the averaging efficiency of such experiments should go up, increasing the practicality of single trial paradigms. The use of true single trials to explore certain aspects of behavior is likely to be restricted to the very high field scanners. Perhaps this is not the limitation we make it out to be since the number of scanners at 3 T and above has tripled in the 1990s and is approaching three dozen. Even more dedicated fMRI centers will be established, as cognitive neuroscientists seek independence from scanners in radiology departments.

A similar thought process may be used with regard to spatial resolution. Sub-millimeter spatial resolution is hardly required for most cognitive experiments. Large areas of the cortex are involved in even the simplest of human behaviors and isotropic 3 mm voxels may be ideal for most experiments. Such volumes are achievable on current clinical hardware and have been exploited to beautiful effect in functional mapping of the visual cortex[9]. Subtleties of cortical functional architecture will likely remain the domain of very high field scanners, but the limits may not be much higher than those probed here for the following reasons.

There are many technical impediments to pushing the spatial and temporal resolution further. Primary among these are the limited signal-to-noise ratio of the image (particularly as images are acquired faster) as well as noise sources intrinsic to the measurement process. Cortical signal changes in fMRI of common stimulation tasks are small (<3%) even at very high magnetic field strengths. Although the instrumental noise in modern MRI instruments can be well below 1% peak-to-peak in a tissue-mimicking object, the signal fluctuation observed in the brain can be much larger. Even when motion is minimized by restraining the head, there are still signal fluctuations of physiological origin amongst which cardiac and respiratory induced components are the most dominant. With the fastest imaging techniques that can make an image in ~50 ms, effectively freezing some types of motion, the image-to-image signal fluctuation can still exceed several percent. When physiological monitoring signals are collected during MRI measurements, some retrospective signal correction can remove the signal fluctuation induced by cardiac 'noise' and respiration[35] to a useful degree. When MRI image acquisition is carried out at double (or more) the highest frequency of interest, the physiological fluctuations can be visualized in the power spectrum of the time series data[36,37]. Such physiological oscillations present in the time series can be filtered out in the post processing[37,38]. Even if these issues can be conquered, there remain a host of biological variables. For example, it is clear that the vascular supply is not regulated on the scale of individual neurons, and might in fact be limited to 0.5 to 1.5 mm in humans depending on field strength[39,40]. If true, this would limit the ultimate spatial resolution achievable with fMRI. Similarly, it is likely that the neural-hemodynamic coupling constant varies between brain areas and even generally between people. Correlations of the type described in this article may allow relative timing between areas to be determined, but the absolute value of those numbers currently seems out of reach. Thus nature, rather than our technology, may set the ultimate limit on how far we may push the limits of fMRI in the future.

Outstanding questions

- Does the fMRI response correspond to synaptic activity or action potentials? The answer has implications in the interpretation of the sign of the fMRI signal change.
- How does the variability of the hemodynamic response in different brain areas and different people at different ages limit the applicability of fMRI chronometry? Averaging across subjects will be difficult if the baseline response functions are different.
- Is the hemodynamic response measured with BOLD confined to the site of spiking or synaptic activity? Ultimately this will determine the spatial resolution and accuracy of the fMRI technique.

Acknowledgements

The authors wish to acknowledge the support of the Medical Research Council of Canada Operating and Salary Support (to R.M.), NIH Grant EY11551 (to R.M.), McDonnell-Pew Progam in Cognitive Neuroscience (to R.M.), and NIH Grants RR08079 and MH57180 (to S-G.K).

References

1 Donders, F.C. (1969) On the speed of mental processes (translation) *Acta Psychol.* 30, 412–431

2 Posner, M.I. (1978) *Chronometric Explorations of Mind*, Oxford University Press

3 Georgopoulos, A.P., Taira, M. and Lukashin, A. (1993) Cognitive neurophysiology of the motor cortex *Science* 260, 47–52

4 Georgopoulos, A.P. and Pellizzer, G. (1995) The mental and the neural: Psychological and neural studies of mental rotation and memory scanning *Neuropsychologia* 33, 1531–1547

5 Alivisatos, B. and Petrides, M. (1997) Functional activation of the human brain during mental rotation *Neuropsychologia* 35, 111–118

6 Cattell, J.M. (1886) The influence of the intensity of the stimulus on the length of the RT *Brain* 8, 512–515

7 Churchland, P.S. and Sejnowski, T.J. (1988) Perspectives on cognitive neuroscience *Science* 24, 741–745

8 Sereno, M.I. *et al.* (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging *Science* 268, 889–893

9 Tootell, R.B.H. *et al.* (1996) New images from human visual cortex *Trends Neurosci.* 18, 481–489

10 DeYoe, E.A. *et al.* (1996) Mapping striate and extrastriate visual areas in human cerebral cortex *Proc. Natl. Acad. Sci. U. S. A.* 93, 2382–2386

11 Tootell, R.B.H. *et al.* (1998) From retinotopy to recognition: fMRI in human visual cortex *Trends Cognit. Sci.* 2, 174–183

12 Blamire, A.M. *et al.* (1992) Dynamic mapping of the human visual cortex by high-speed magnetic resonance imaging *Proc. Natl. Acad. Sci. U. S. A.* 89, 11069–11073

13 Savoy, R.L. *et al.* (1994). Exploring the temporal boundaries of fMRI: Measuring responses to very brief visual stimuli *Soc. Neurosci. Abstr.* 20, 1264

14 Menon, R.S. *et al.* (1995) BOLD-based functional MRI at 4 Tesla includes a capillary bed contribution: echo-planar imaging correlates with previous optical imaging using intrinsic signals *Magn. Reson. Med.* 33, 453–459

15 Buckner, R.L. *et al.* (1996) Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging *Proc. Natl. Acad. Sci. U. S. A.* 93, 14878–14883

16 Richter, W. *et al.* (1997) Sequential activity in human motor areas during a delayed cued finger movement task studied by time-resolved fMRI *NeuroReport* 8, 1257–1261

17 Rosen, B.R., Buckner, R.L. and Dale, A.M. (1998) Event-related functional MRI: Past, present and future *Proc. Natl. Acad. Sci. U. S. A.* 95, 773–780

18 Menon, R.S., Luknowsky, D.L. and Gati, J.S. (1998) Mental chronometry using latency-resolved functional magnetic resonance imaging *Proc. Natl. Acad. Sci. U. S. A.* 95, 10902–10907

19 Kim, S-G., Richter, W. and Ugurbil, K. (1997) Limitations of temporal resolution in functional MRI *Magn. Reson. Med*. 37, 631–636

20 Richter, W. *et al.* (1997) Time-resolved fMRI of mental rotation *NeuroReport* 8, 3697–3702

21 Toni, I. *et al.* (1999) Signal-, set- and movement-related activity in the human brain: an event-related fMRI study *Cereb. Cortex* 9, 35–49

22 Lee, K.M., Chang, K.H. and Roh, J.K. (1999) Subregions within the supplementary motor area activated at different stages of movement preparation and execution *NeuroImage* 9, 117–123

23 Roy, C.S. and Sherrington, C.S. (1890) On the regulation of blood supply of the brain *J. Physiol.* 11, 85–108

24 Fox, P.T. *et al.* (1986) Mapping human visual cortex with positron emission tomography *Nature* 323, 806–809

25 Friston, K.J. (1997) Imaging cognitive anatomy *Trends Cognit. Sci.* 1, 21–27

26 Rabe-Hesketh, S., Bulmore, E.T. and Brammer, M.J. (1997) The analysis of functional magnetic resonance images *Stat. Methods Med. Res.* 6, 215–237

27 Ogawa, S. *et al.* (1992) Intrinsic signal changes accompanying sensory stimulation: functional brain mapping using MRI *Proc. Natl. Acad. Sci. U. S. A.* 89, 5951–5955

28 Kwong, K.K. *et al.* (1992) Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation *Proc. Natl. Acad. Sci. U. S. A.* 89, 5675–5679

29 Bandettini, P.A. *et al.* (1992) Timecourse EPI of human brain function during task activation *Magn. Reson. Med*. 25, 390–398

30 Shepard, R.N. and Metzler, J. (1971) Mental rotation of three-dimensional objects *Science* 171, 701–703

31 Alexander, G.E. and Crutcher, M.D. (1990) Neural representations of the target (goal) of visually guided arm movements in three motor areas of the monkey *J. Neurophysiol.* 64, 164–178

32 Wise, S.P., Weinrich, M. and Mauritz, K-H. (1986) *Prog. Brain Res.* 64, 117–131

33 Zarahn, E., Aguirre, G.K. and D'Esposito, M. (1999) Temporal isolation of the neural correlates of spatial mnemonic processing with fMRI, *Brain Res. Cogn. Brain Res.* 7, 255–268

34 Samuel, M. *et al.* (1998) Exploring the temporal nature of hemodynamic responses of cortical motor areas using functional MRI *Neurology* 51, 1567–1575

35 Hu, X. *et al.* (1995) Retrospective estimation and compensation of physiological fluctuation in functional MRI *Magn. Reson. Med.* 34, 210–221

36 Mitra, P.P. and Pesaran, B. (1999) Analysis of dynamic brain imaging data *Biophys. J.* 76, 691–708

37 Thomas, C.G. and Menon, R.S. (1998) Amplitude response and stimulus presentation frequency response of human primary visual cortex using BOLD EPI at 4 T *Magn. Reson. Med.* 40, 203–209

38 Biswal, B. *et al.* (1996) Reduction of physiological fluctuations in fMRI using digital filters *Magn. Reson. Med.* 35, 107–113

39 Menon, R.S. and Goodyear, B.G. (1999) Submillimeter functional localization in human striate cortex using BOLD contrast at 4 Tesla: implications for the vascular point spread function *Magn. Reson. Med.* 41, 230–235

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

Autism: cognitive deficit or cognitive style?

Francesca Happé

Autism is a developmental disorder characterized by impaired social and communicative development, and restricted interests and activities. This article will argue that we can discover more about developmental disorders such as autism through demonstrations of task success than through examples of task failure. Even in exploring and explaining what people with autism find difficult, such as social interaction, demonstration of competence on contrasting tasks has been crucial to defining the nature of the specific deficit. Deficit accounts of autism cannot explain, however, the assets seen in this disorder; for example, savant skills in maths, music and drawing, and islets of ability in visuospatial tests and rote memory. An alternative account, reviewed here, suggests that autism is characterized by a cognitive style biased towards local rather than global information processing – termed 'weak central coherence'. Evidence that weak coherence might also characterize the relatives of people with autism, and form part of the extended phenotype of this largely genetic disorder, is discussed. This review concludes by considering some outstanding questions concerning the specific cognitive mechanism for coherence and the neural basis of individual differences in this aspect of information processing.

Autism is a devastating developmental disorder affecting at least one in a thousand children and adults. Although biologically based, with a strong genetic component, diagnosis of autism is still made by behavioural criteria: qualitative impairments in social and communicative development, with restricted and repetitive activities and interests[1]. It is not difficult to find things that people with autism have difficulty with – indeed, most autistic people also have general learning difficulties and low IQ. However, I will argue in this review that progress in understanding this disorder, and its implications for normal development, will arise chiefly through exploration of what people with autism are *good* at.

Understanding preserved and impaired abilities in autism

Much progress has been made in the last 15 years in understanding the nature of the social and communicative handicaps in autism. Primary in this has been the notion that people with autism fail to represent the mental states of others (and possibly of self) – a deficit in what has been called 'theory of mind' (see Box 1). This account can explain why children with autism have such difficulty with simple behaviours such as joint attention, pretend play and even telling lies[2]. However, these deficits, and failure on key tasks such as false-belief tests, are only informative when viewed against a background of task success. Clearly, (behavioural) task failure is ambiguous with regard to underlying (cognitive) deficits; a child might fail a test for any number of uninteresting reasons, such as lack of motivation, attention or task comprehension. To isolate the reason for task failure and to rule out alternative explanations, closely-matched control tasks have been used. So, for example, the autistic failure to understand deception (manipulating beliefs) is interesting only when contrasted with success on control tasks involving sabotage (manipulating behaviour)[2]. This research, showing preserved as well as deficient social skills, has clarified the nature of the social impairment in autism;

F. Happé is at the Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, 111 Denmark Hill, London, UK SE5 8AF.

tel: +44 171 919 3873
fax: +44 171 9191 3866
e-mail: f.happe@iop.kcl.ac.uk

1364-6613/99/$ – see front matter PII: S1364-6613(99)01318-2