PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Event-related functional magnetic resonance imaging: modelling, inference and optimization

Oliver Josephs and Richard N. A. Henson

Phil. Trans. R. Soc. Lond. B 1999 **354**, 1215-1228
doi: 10.1098/rstb.1999.0475

THE ROYAL SOCIETY

Event-related functional magnetic resonance imaging: modelling, inference and optimization

Oliver Josephs[1*] **and Richard N. A. Henson**[1,2]

[1]*Wellcome Department of Cognitive Neurology, Institute of Neurology, 12 Queen Square, London WC1N 3BG, UK*
[2]*Institute of Cognitive Neuroscience, University College London, 10 Queen Square, London WC1N 3BG, UK*

Event-related functional magnetic resonance imaging is a recent and popular technique for detecting haemodynamic responses to brief stimuli or events. However, the design of event-related experiments requires careful consideration of numerous issues of measurement, modelling and inference. Here we review these issues, with particular emphasis on the use of basis functions within a general linear modelling framework to model and make inferences about the haemodynamic response. With these models in mind, we then consider how the properties of functional magnetic resonance imaging data determine the optimal experimental design for a specific hypothesis, in terms of stimulus ordering and interstimulus interval. Finally, we illustrate various event-related models with examples from recent studies.

Keywords: single trial; event-related fMRI; statistics; echo-planar MRI; BOLD contrast

1. INTRODUCTION

The term event-related functional magnetic resonance imaging (efMRI) can be defined as the use of functional magnetic resonance imaging (fMRI) to characterize and detect transient haemodynamic responses to brief stimuli or tasks. Event-related, or trial-based, measurement is already standard in the field of electrophysiology, namely stimulus-locked, event-related potentials (ERPs). In contrast, efMRI is a relatively new technique. fMRI measurements of the haemodynamic response are as recent as those of DeYoe *et al*. (1992), and statistical methods for detecting event-related responses are as recent as those of Boynton *et al*. (1996). The method has since proved popular: more than 91 published papers have used the technique.

Here we review the technique. The stress is on the statistical modelling techniques used to detect reliable event-related responses. It is hoped that this will complement other review articles by Rosen *et al*. (1998) and D'Esposito *et al*. (1999), which concentrate on the nature of the event-related response and give examples of applications in cognitive neuroscience. However, we shall recapitulate the advantages of the event-related approach over previous methods (§ 2), and briefly outline the basics of haemodynamic responses and their measurement (§ 3). We shall then discuss the reliable detection of event-related fMRI responses, which involves creating models (§ 4) and using the models to make inferences (§ 5). We shall then introduce two areas of current research: the applications of random effect models in single subject efMRI and the optimization of event-related experimental designs (§ 6). Finally, we shall present several recent neuroscientific applications that illustrate the modelling techniques (§ 7).

*Author for correspondence (o.josephs@fil.ion.ucl.ac.uk).

2. THE IMPORTANCE OF BEING EVENT RELATED

Previous methods of functional neuroimaging, such as positron emission tomography (PET), have limited temporal resolution, which precludes the imaging of all but prolonged states of brain activity. Such state-based designs were initially adopted for fMRI studies. However, it did not take long before the greater sensitivity and temporal resolution of fMRI, particularly at higher magnetic field strengths, were used to allow the more flexible, event-related approach.

(a) *Trial-based designs*

The cardinal advantage of efMRI over state-related fMRI is that it allows trial-based rather than block-based experiments. Trial-based experiments offer several advantages.

(i) The order of trials can be randomized, meaning that the response to a trial is neither confounded by a subject's cognitive set nor systematically influenced by previous trials (Johnson *et al*. (1997), for example, showed that ERPs differed depending on whether stimuli were randomized or blocked).
(ii) Trials can be individually categorized or parametrized post hoc according to a subject's performance, as indexed by accuracy or reaction time for example (see, for example, Wagner *et al*. 1998).
(iii) Some experiments involve events that cannot be blocked, such as 'oddball' paradigms, in which the event of interest is a stimulus that violates the prevailing context (B. A. Strange, R. N. A. Henson, K. J. Friston and R. J. Dolan, unpublished data).
(iv) Some events can occur unpredictably, and can be indicated only by the subject (such as spontaneous transitions in the perception of ambiguous figures (see, for example, Kleinschmidt *et al*. 1998).

(v) Trial-based fMRI results are more directly comparable with other trial-based neuroscientific methods (for example, ERPs and reaction-time measurements).

(b) *Block-based designs*

Not only do event-related techniques make trial-based experiments feasible but, more generally, they provide improved methods for analysing other forms of fMRI experiment. For example, from an event-related perspective, a state can be modelled, to first order, as a continuous train of events, each representing one trial within the block. If the trials are relatively sparse within the block, the event-related approach can account for fluctuation of the response within the block. This means that the experimental variance is likely to be better modelled than with a simple state-based approach (Price 1999). Another example is when a state has variance associated with the event of entering the state, an early effect (Friston *et al*. 1995*a*). This can simply be treated by modelling an additional event representing the beginning of the block in addition to the state-related variance.

3. MEASURING EVENT-RELATED RESPONSES

In this section, we briefly examine the nature of the event-related haemodynamic response and its measurement using echo-planar magnetic resonance imaging (EPI). The fMRI signal is caused by the (sluggish) haemodynamic response after neuronal activation. The biophysical mechanisms are not fully understood, but the signal change is clearly a function of blood flow, volume and oxygenation state (Howseman & Bowtell, this issue). This blood oxygenation level dependent (BOLD) contrast is usually measured with T_2^*-weighted gradient-recalled EPI.

(a) *Variability in the response*

The question arises of whether the BOLD response is a reliable and consistent marker of neuronal activity.

(i) *Temporal*

Aguirre *et al*. (1998) examined the variability of the BOLD response in the central sulcus during a simple reaction-time task performed across scanning sessions within a day, between days and between subjects. Variability across sessions within a day was significant for only one subject. Significant variability across days was observed in three out of four subjects (although, because scanning across days entailed repositioning subjects in the scanner, it was not possible to attribute this variability to temporal fluctuations in BOLD response itself or to differences in the scanning environment). Reproducibility of state-related responses across multiple sessions in several regions of interest has been demonstrated (McGonigle 1999), although other brain regions showed significant session–state interactions. These studies highlight the risk of false-positive results when generalizing from single sessions.

(ii) *Spatial*

Although a superficial comparison of responses measured in motor cortex (see, for example, Aguirre *et al*. 1998), visual cortex (see, for example, Boynton *et al*. 1996) and auditory cortex (see, for example, Josephs *et al*. 1997) suggests that the basic form of the BOLD response is quite typical across the brain, one might expect variability arising from differences in the blood supply to different regions and the extent to which different components of the venous compartment contribute to the MRI signal. However, there are surprisingly few quantitative studies of this spatial variability. Schacter *et al*. (1997) found that responses in anterior prefrontal cortex were best fitted by gamma functions that peaked 4 s later than the functions that best fitted the responses in the visual cortex. Although the vasculature of the two regions does not differ markedly, it remains uncertain whether this relatively delayed prefrontal BOLD response reflects haemodynamic or neuronal effects. Clarification of this issue requires the mapping of venous flow in different brain regions (see, for example, Lee *et al*. 1995).

(iii) *Individual*

By far the largest source of variability observed by Aguirre *et al*. (1998) was across subjects, presumably reflecting considerable individual differences in physiology. Physiological manipulations have been shown to offset the BOLD response (Corfield *et al*. 1998), although in this case they did not modulate the magnitude, relative to the baseline, of responses evoked visually.

Thus preliminary studies suggest that the BOLD response shows reasonable reproducibility across sessions and brain regions within a subject, but significant variability across subjects. Further work is clearly needed: for example, to quantify the extent to which this variability affects the appropriateness of statistical models that assume a single canonical form for the event-related response. Models that use more general basis sets (§ 4(c)) are better able to accommodate variability in the BOLD response.

(b) *Linearity*

The assumption that the magnitude of the BOLD response is linearly related to the magnitude of underlying neuronal activity is useful in simplifying the modelling and detection of event-related responses. Various studies (Boynton *et al*. 1996; Dale & Buckner 1997; Pollmann *et al*. 1998) have concluded that a reasonable degree of linearity is to be expected, although this is clearly not always true (Vasquez & Noll 1998), particularly for short interstimulus intervals (Friston *et al*. 1998*b*). Work is needed to correlate these effects with electrophysiological investigations to discover which nonlinear effects occur at the neuronal level and which occur at the haemodynamic level. Some manifestations of nonlinearity might include the following.

(i) *Saturation*

Whereas the PET response seems to increase linearly with stimulus intensity, the fMRI response seems to saturate at high levels (Howseman & Bowtell, this issue).

(ii) *Habituation*

The response to a trial might depend on the history of previous trials, over and above the linear superposition of previous responses. ERPs to rapid tone presentations, for example, are attenuated at short interstimulus intervals,

presumably owing to habituation of the cells in the auditory cortex. This form of habituation is also likely to be reflected by the haemodynamic responses measured in efMRI.

(iii) *Slew limits*

Slew limits are effectively the maximum rate at which the response can change. It might be that the haemodynamic system can reach a final level that depends linearly on the applied stimulus, but can only reach that level at a certain rate.

From the perspective of modelling, however, we can remain agonistic on the question of linearity. We shall demonstrate in § 4 that, even if the linear assumptions are violated, we simply expand our models to include nonlinear terms and proceed to make statistical inferences on the basis of the linear or nonlinear component of the response.

(c) *Whole-brain EPI measurement*

fMRI data are usually acquired by using EPI, which is fast relative to other MRI imaging procedures: a single slice can be measured in less than 100 ms. For single-slice studies this temporal resolution is sufficient to sample event-related haemodynamic responses, the spectral power density of which falls rapidly above 0.1 Hz (see figure 1*a*). However, a multi-slice scan of the whole brain at reasonable spatial resolution, which is necessary for many experiments, entails a volume repeat time (TR) of several seconds, depending on the number of slices. Thus the measurement of a haemodynamic response can often undersample the experimental variance. One method to overcome this problem is to stagger stimulus times relative to scan acquisition times, realizing a higher effective sampling rate (Josephs *et al*. 1997).

(d) *Noise sources*

Event-related responses are not recorded in isolation, but rather in the context of various non-deterministic confounding effects. These include instrument noise, pulsatile physiological noise due to cardiac and respiratory rhythms, endogenous haemodynamic fluctuations and residual effects of subject movement (motion effects can persist even after image realignment, because of the non-rigid body nature of the multi-slice imaging). These noise sources dominate at low frequencies, having a $1/f$ form (Aguirre *et al*. 1997; Zarahn *et al*. 1997*b*). The noise can be removed by high-pass filtering during preprocessing or modelling (§ 5(b)) (Holmes *et al*. 1997).

4. MODELLING EVENT-RELATED RESPONSES

In § 3 we summarized the nature of the haemodynamic response to brief stimuli and the means by which we can measure it with MRI. In this section we explain how, armed with this knowledge, we can predict the forms of variance that we expect to measure in an event-related experiment.

Several early event-related studies did not model the BOLD response explicitly, but implicitly allowed for haemodynamic lag by equating the event-related response to the signal measured in scans acquired 4–6 s after the stimulus (McCarthy *et al*. 1996). By synchronizing stimuli with scan onsets and employing long interstimulus intervals, time-courses can be plotted over peristimulus scans (see, for example, Cohen *et al*. 1997). However, a more powerful technique is to model the predicted time-course of the fMRI signal (Friston *et al*. 1995*a*; Josephs *et al*. 1997). Below, we describe the generation of a model from component models of the stimulus, neuronal activity and haemodynamic response.

(a) *The general linear model*

We formulate models within a general linear model (GLM) framework. By following this approach we can vary the complexity of the model to include higher-order effects such as differential activation, interactions with state or time, and even (linear approximations to) nonlinear effects (§ 5(f)). In addition, we can account for the temporal covariance structure of fMRI data.

In a GLM, we assume that the observed data can be expressed as a linear combination of explanatory variables plus an error term. In matrix notation,

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{y}$ is a vector of the measured results (a time-series sampled every scan), X is the design matrix, in which each column represents one explanatory time-course (covariate), $\boldsymbol{\beta}$ is a vector of weights or parameters for each covariate, and $\boldsymbol{e}$ is a residual error term with zero mean and covariance V. A least-squares solution, $\boldsymbol{\beta}'$ for $\boldsymbol{\beta}$ in such a model is

$$\boldsymbol{\beta}' = \mathrm{pinv}(X^{\mathrm{T}}\mathrm{inv}(V)X)X^{\mathrm{T}}\mathrm{inv}(V)\boldsymbol{y},$$

which, for independent, identically distributed residuals, reduces to

$$\boldsymbol{\beta}' = \mathrm{pinv}(X^{\mathrm{T}}X)X^{\mathrm{T}}\boldsymbol{y}.$$

Thus an important assumption in the analysis of time-series is that the residuals are white, that is they depend only on the number of degrees of freedom (in fact, the noise is often coloured owing to serial autocorrelations in the time-series, which requires corrections to the degrees of freedom; see Worsley & Friston (1995) for a complete description). Under the null hypothesis, the model is also expected to model variance depending only on the number of degrees of freedom (covariates) in the model (§ 5(d)).

(b) *Stimulus model*

We assume that the stimulus or task can be characterized by a multivariate function of time, $\boldsymbol{s}(t)$, that has one dimension per experimental factor 1, . . ., n. Values of $\boldsymbol{s}(t) = 0$ denote the baseline (no stimulus being applied). The occurrence of a brief stimulus or task is modelled as a Kroneker delta function. This model can also account for state changes (e.g. a block of stimulation) with appropriate 'top-hat' functions. The resolved magnitude of $\boldsymbol{s}(t)$ can be set arbitrarily to 1, to act as an indicator representing the presence of a categorical factor, or to a value representing the instantaneous magnitude of a parametric factor.

(c) *Neuronal activation model*

The neuronal activation to stimulation, $u(t)$, can be expressed as a function of $\boldsymbol{s}(t)$:

$$u(t) = \phi\{\boldsymbol{s}(t)\} + e_{\mathrm{s}}(t),$$

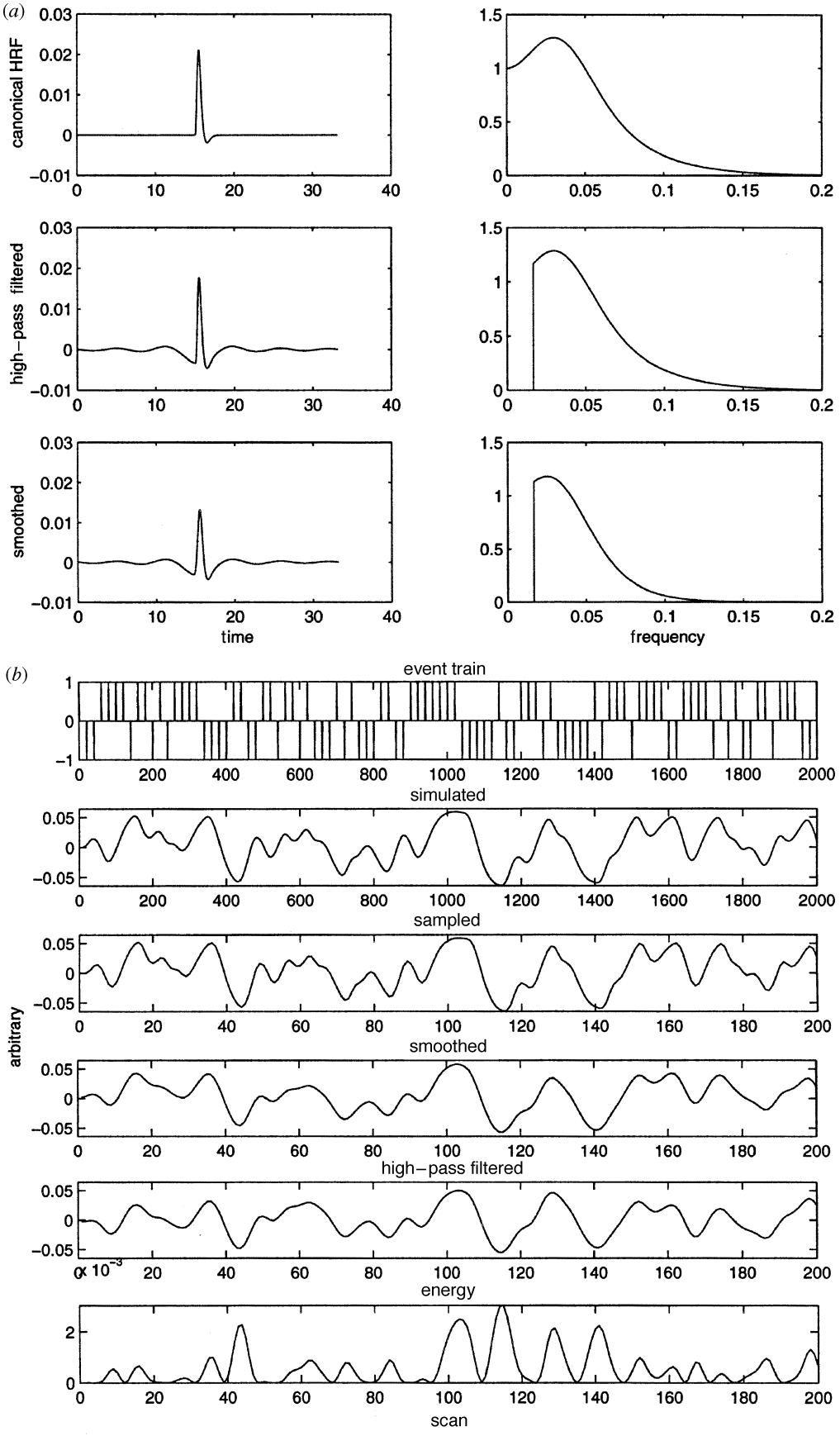

Figure 1. (*a*) EHRF in time (s) and frequency (Hz) domains after temporal smoothing with a FWHM Gaussian kernel of 4 s and filtering with an ideal high-pass cut-off of 1/60 s. (*b*) Sequence from a simulation of the predicted signal energy per scan (bottom plot) after convolving a differential stimulus function (top plot) with a canonical HRF at a resolution of 0.1 s (second plot), sampling every TR of 1 s (third plot), smoothing with a FWHM Gaussian kernel of 4 s (fourth plot), and high-pass filtering to 1/60 s (fifth plot). The smoothing and filtering stages are separated (rather than being convolved with the EHRF in (*a*)) to illustrate their effects separately.

where $e_s(t)$ represents spontaneous neuronal activity. In general, the function ϕ is nonlinear but we can expand it as a power series,

$$u(t) = \boldsymbol{s}^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{s}^{\mathrm{T}}B\boldsymbol{s} + \ldots + e_s(t), \quad (2)$$

where $\boldsymbol{b}$ is a vector of n linear coefficients and the symmetrical $n \times n$ matrix B contains quadratic coefficients. The diagonal components of B represent nonlinear influences of a given factor on neuronal activity (e.g. saturation of receptors at high levels of stimulation). The off-diagonal terms encode interactions between factors (e.g. $B_{12} = B_{21}$ represents an interaction between factor 1 and factor 2).

(d) *Haemodynamic response model*

We know that the haemodynamic response caused by neuronal activation lasts for a finite duration after the event. This is equivalent to observing that a given measurement is influenced only by neuronal activation that occurred within an equal, finite period before the measurement. This restricted temporal influence on a scan can be expressed by a finite memory model. With the use of a discrete-time formulation (for notational simplicity), the finite memory model can be expressed as

$$y(t) = \theta\{\boldsymbol{u}(t)\} + e_m(t),$$

where the T-dimensional vector $\boldsymbol{u}(t) = [u(t), u(t-1) \ldots u\{t-(T-1)\}]$ represents the neuronal activation at time t to a memory depth of T, and $e_m(t)$ represents measurement noise at time t. In general, θ is a nonlinear function but it can be expanded as a Volterra series (Friston *et al.* 1998*b*),

$$y(t) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{h} + \boldsymbol{u}^{\mathrm{T}}H\boldsymbol{u} + \ldots + e_m(t), \quad (3)$$

where $\boldsymbol{h}$ is the linear convolution term and H is a $T \times T$ matrix of second-order coefficients. The nonlinearity coded by H includes the saturation of the BOLD response to neuronal activation.

We assume that $\boldsymbol{h}$ can be expressed as a linear combination of $i = 1, \ldots p$ temporal basis functions, $g_i(t)$. In discrete-time matrix form,

$$\boldsymbol{h} = a_1 g_1 + a_2 g_2 + \ldots + a_p g_p.$$

In a similar manner, we assume that H can be expressed as the linear combination of $j = 1 \ldots q$ second-order basis functions, $G_j(t)$:

$$H = A_1 G_1 + A_2 G_2 + \ldots + A_q G_q.$$

In the general case, we cannot separately estimate the neuronal and haemodynamic parameters implicit in equation (3), given only the stimulus model as input and the data as output. For example, we cannot distinguish between saturation at the neuronal level (the on-diagonal terms of B in equation (2)) and saturation at the haemodynamic level (the terms of H in equation (3)). Nonetheless, other parameters (such as the off-diagonal terms of B that capture interactions between types of event) can be estimated, at least to a proportional level, which suffices for inferences by means of a variance ratio (F-test; § 5(c)). Moreover, this general framework permits the derivation of simpler models by following additional assumptions, such as linearity at the neuronal or haemodynamic levels, as illustrated by the examples in § 4(f).

(e) *Measurement model*

The high-resolution discrete-time model in equation (3) is sampled every TR, such that the predicted MRI signal is $y(mT_R)$, where T_R is the scan repetition time and m is the scan number. The model (and data) can also be smoothed temporally to swamp intrinsic autocorrelation in the data with a known autocorrelation (Worsley & Friston 1995).

(f) *Examples*

We now consider two special cases of the above model and demonstrate the GLM form for the measured time-course that ensues.

(i) *Univariate indicator stimulus function*

Here we assume that only one stimulus type is presented ($n = 1$), its intensity is constant, and neuronally generated noise is negligible ($e_s(t) = 0$). With these simplifications there are no interactions between event types and no nonlinearity in the neuronal activity, because only two values of neuronal activity are generated. Thus the neuronal activation model is

$$u(t) = bs(t)$$

and the recent history of the neuronal activation is

$$\boldsymbol{u}(t) = b\boldsymbol{s}(t),$$

where $\boldsymbol{s}(t)$ is formed by analogy with $\boldsymbol{u}(t)$ above. This can be substituted into the haemodynamic response model,

$$y(t) = b\boldsymbol{s}^{\mathrm{T}}\boldsymbol{h} + b^2\boldsymbol{s}^{\mathrm{T}}H\boldsymbol{s} + \ldots + e_m(t),$$

and substituting for $\boldsymbol{h}$ and H yields

$$y(t) = a_1 b\boldsymbol{s}^{\mathrm{T}}\boldsymbol{g}_1 + a_2 b\boldsymbol{s}^{\mathrm{T}}\boldsymbol{g}_2 + \ldots + a_p b\boldsymbol{s}^{\mathrm{T}}\boldsymbol{g}_p + A_1 b^2\boldsymbol{s}^{\mathrm{T}}G_1\boldsymbol{s} + A_2 b^2\boldsymbol{s}^{\mathrm{T}}G_2\boldsymbol{s} + \ldots + A_q b^2\boldsymbol{s}^{\mathrm{T}}G_q\boldsymbol{s} + e_m(t).$$

This is a GLM for y with $(p+q)$ unknown parameters comprising neuronal and haemodynamic variables of the known stimulus function (event train) and known temporal basis functions (compare equation (1)).

(ii) *Linear haemodynamic response*

We now consider the case for n stimulus factors and for which the haemodynamic response is linear ($H = 0$). For simplicity we again neglect the neuronal noise term, $e_s(t)$. In this case, $\boldsymbol{u}(t)$ is a vector containing T elements, each of which is the sum of $n + \frac{1}{2}n(n+1)$ terms:

$$\boldsymbol{u}(t) = S^{\mathrm{T}}\boldsymbol{b} + S^{\mathrm{T}}B\boldsymbol{s},$$

where S is an $N \times T$ matrix, and thus

$$y(t) = a_1 b_1 \boldsymbol{s}_1^{\mathrm{T}}\boldsymbol{g}_1 + a_1 b_2 \boldsymbol{s}_2^{\mathrm{T}}\boldsymbol{g}_1 + \ldots + a_2 b_1 \boldsymbol{s}_1^{\mathrm{T}}\boldsymbol{g}_2 + \ldots + a_p b_n \boldsymbol{s}_n^{\mathrm{T}}\boldsymbol{g}_p + \ldots + a_1 B_{11} s_1(t)\boldsymbol{s}_1^{\mathrm{T}}\boldsymbol{g}_1 + a_1 B_{21} s_2(t)\boldsymbol{s}_1^{\mathrm{T}}\boldsymbol{g}_1 + \ldots + a_2 B_{11} s_1(t)\boldsymbol{s}_1^{\mathrm{T}}\boldsymbol{g}_2 + \ldots + a_p B_{nn} s_n(t)\boldsymbol{s}_n^{\mathrm{T}}\boldsymbol{g}_p + e_m(t).$$

The resulting GLM has $p\{n + \frac{1}{2}n(n+1)\} = pn\frac{1}{2}(n+3)$ unknown parameters. Thus in a two-factor experiment modelled by three linear basis functions, the model would contain 15 covariates: six to model the two main effects,

six to model nonlinearity in the factors, and three to model the interaction.

(g) *Haemodynamic response function*

We now turn to examine the choice of basis functions, $\boldsymbol{g}_i(t)$. These can be either direct expansions of time or expansions of a prior estimate of a typical haemodynamic response function (HRF). An example of the former approach is the use of Fourier series components to estimate the full effective sampling-rate-limited component of the response (Josephs *et al*. 1997). For the latter approach, we typically employ a multivariate Taylor expansion of a gamma-function approximation to a canonical HRF (Friston *et al*. 1998*a*). The higher-order basis functions in this expansion include the partial derivatives of the HRF with respect to time and dispersion. The Fourier basis set has the advantage of generality and insensitivity to slice timing (§4(h)), whereas the partial derivatives of a canonical HRF have the advantage that the parameter for each covariate is interpretable in terms of response magnitude, latency or duration (§5(f)(iv) and §7(f)). The canonical form and its derivatives can be tested separately by means of univariate tests, whereas the Fourier components are best tested by reduced *F*-tests (§5(d)).

(h) *Slice-timing issues*

A single model is often assumed for all voxels (i.e. scan-based or volume-based design matrices). At present, however, whole-brain EPI acquisitions (§3(c)) require a time comparable to the haemodynamic rise time, meaning that different slices are acquired at different times relative to the response. This can be partly corrected for by temporal interpolation. However, interpolation is not ideal, particularly if significant experimental power exists at frequencies above the Nyquist sampling limit $(2\mathrm{TR})^{-1}$. Alternatively, slice-timing differences can be accommodated by a suitable choice of basis functions, such as the phase-invariant Fourier set, or the inclusion of temporal derivatives (Henson *et al*. 1999). However, the most general solution is to construct separate models for each slice that take into account their acquisition time relative to the stimulus (i.e. slice-based design matrices).

5. INFERENCE IN efMRI

Having constructed a model for the event-related variance, we now discuss the estimation of the parameters required to fit the model to the data and methods for performing statistical inference with the use of the parameter and residual variance estimates. The questions that can be addressed mirror the models that have been used. We consider only inferences at the level of individual voxels (for issues regarding whole-brain inferences, see Friston *et al*. (1995*b*)). Inference involves the use of residual variance estimates to determine the distribution of the parameter estimates for effects of interest under the null hypothesis, and hence the probability of falsely rejecting the null hypothesis (the significance), assuming the null hypothesis to be true.

We could, in principle, test for activation by following a single event-related trial. However, this is unlikely to suffice because the event-related response is contaminated by at least two sources of error. First, instrument noise is typically quite high (although much lower than in PET). Ratios of 2 or 3 for signal to (r.m.s.) noise are typical, and so any activation will only be detected with a low significance. Higher field systems, with higher sample magnetization, might allow the detection of the event-related response to single trials (Richter *et al*. 1997). Second, even though the sensitivity to detect the haemodynamic response is high, the underlying brain response itself often has a random component. One reason is that the brain region in question might have functions unrelated to the task under investigation. Furthermore, subjects might vary their strategy from trial to trial, yielding a random, experimentally induced variance component. For this reason one wishes to detect a typical rather than a single response. The question therefore becomes 'Which voxels consistently show an event-related response?' However, we begin by discussing the concepts of confounds, parameter estimation, and contrasts.

(a) *Confounds*

We routinely model several confounding effects when analysing fMRI data. These effects include the mean over time, all low frequencies up to some cut-off point (typically 60 s), to remove low-frequency noise (§3(d)), and a global mean effect over space (i.e. all voxels). The important point with respect to efMRI is that the experimental variance is predominantly of a relatively high frequency compared with a state-related design. This allows us to apply a higher cut-off to the high-pass filter, removing more of the low-frequency effects that would otherwise reduce sensitivity.

(b) *Estimation of parameters and residual variance*

Our statistical inference uses an analysis of covariance procedure. We need to discover the parameter estimates for the linear components of the full model and, in addition, the residual variance accounting for (i) the confounds only (the reduced model (Buechel *et al*. 1996)), and (ii) both the effects of interest and the confounds. The purpose of the reduced model is to estimate the residual variance and degrees of freedom when accounting only for the confounding effects, which can be used in an *F*-test of significance (§5(d)).

(c) *Contrasts*

In many cases we wish to test specific, *a priori* hypotheses about differences between two or more types of event. Such contrasts (Friston *et al*. 1995*b*) are linear combinations of event-related effects, specified as vectors of weights for each event type. The significance of these contrasts can be indexed by *t*-tests (§5(d)). We can incorporate these contrasts into the model by multiplying the design matrix X by a matrix C containing each contrast. This rotation produces a new model, X', in which each column is a linear combination of columns from X, with weightings according to the appropriate row (contrast) from C.

(d) *Fixed-effects inference*

Event-related inference can be performed within the GLM by two types of significance test: Student's *t*-test and Fisher's *F*-test. In both cases we use the results of

fitting the full GLM. For the *F*-test we additionally fit the reduced model (§5(c)).

The *t*-test is used when we wish to determine the significance of a particular contrast. By comparing the parameter estimate against a *t*-distribution, we determine a significance level for the contrast. For the *F*-test we do not refer to the parameter estimates but rather the ratio of error variances obtained from the full and the reduced models. Under the null hypothesis, this ratio is expected to follow an *F*-distribution with $D-N$ and $D-(N-M)$ degrees of freedom, where D is the number of degrees of freedom in the data, N is the degrees of freedom in the complete model and M is the degrees of freedom in the confounds. For independent residual error variance, the degrees of freedom in the data equal the number of scans (for temporally smoothed data, the effective degrees of freedom are less than the number of scans (Worsely & Friston 1995)). For non-collinear covariates, the degrees of freedom are the number of covariates in the model.

The advantages of the *t*-test are that multiple tests can be performed after fitting only one model and that the direction of effects can be tested (e.g. an activation or deactivation). The *F*-test is required when we wish to test for the combined significance of a multiple-component model (e.g. with multiple basis functions). With a single component, the *F*-value obtained in an *F*-test equals the square of the *t*-value from an equivalent *t*-test.

(e) *Inference about a single type of event*

The question asked here is 'Did the voxel respond to the stimulus?' This is the simplest hypothesis, in which we attempt to disprove the null hypothesis that no event-related activation occurred. Under the null hypothesis the variance in the subspace of the model due to the event would be expected to depend only on the degrees of freedom in the model relative to the degrees of freedom in the data. The ratio of these modelled-to-residual variances allows an *F*-test of the significance of the activation. In the context of confounding effects, the *F*-ratio can be formed between the extra variance accounted for by the effects of interest relative to the residuals from the full model (Buechel *et al*. 1996).

In some designs, the response of interest is that to a single event in the context of background events of no interest. An example of this design is an 'oddball' paradigm, in which the event of interest is a stimulus that deviates from surrounding context events. For example, low-pitched tones might be presented rapidly with occasional high-pitched tones. If the context events are presented rapidly enough (relative to the time constants of the HRF), they can be treated as a raised baseline, for which we assume that a tonic haemodynamic state is reached with respect to the common effect (of tones). If the interval between effects of interest (high tones) is sufficiently long, we can detect a significant event-related effect superimposed on this baseline (assuming linearity). If context events are not presented rapidly enough for a tonic state to be reached, they may need to be modelled explicitly as confounds (otherwise the assumption that residual noise is white might be compromised, and type I errors can arise if the associated variance is misattributed to the effect of interest).

(f) *Inference about multiple types of event*

Here we are interested in the differential effects of two or more event types. We construct an $N\times M$ column design matrix of N event types and M basis functions that spans the space of experimental variance. We can now rotate the model (§5(c)) to form a design matrix comprising linear combinations of the original covariates. Under certain useful rotations, this rotated model is equivalent to the original model but the fitted parameter estimates can be more directly interpretable. Note that if events of each type are close together in time and not completely randomized in order (such that the covariates for the different event types are correlated), then some of the experimental variance due to one event type might be inappropriately modelled by the other event types, resulting in a loss of sensitivity to detect responses to each event type (a type II error).

(i) *Differential responses*

An example rotation for two trial types is to form the sum and difference between the two responses to (note that the sum of two event-type covariates, $A+B$, is orthogonal to the difference, $A-B$). These partitions of the model can be interpreted as accounting for the mean (common) activation and the differential activation.

(ii) *Factorial responses*

One can extend the idea of modelling commonalties and differences to more general factorial experiments. For example, the model for a two-factor, two-level experiment comprises four partitions, one for each event type. This model can be rotated to represent, for example, the common effect of a trial, the two main effects and an interaction term.

(iii) *Parametric responses*

The levels of a given factor can be continuous rather than discrete. In this case the model for the response magnitude ($\boldsymbol{s}(t)$ in §4(b)) depends on the level of the factor. Alternatively, the response magnitude might be determined post hoc by behavioural measures (e.g. galvanic skin response) that are associated with each trial.

(iv) *Response parameters*

The components of some basis sets can have different interpretations, and it can therefore be useful to test them separately. The use of a canonical HRF and its derivatives with respect to time and dispersion, for example, allows separate *t*-tests of differences in magnitude, latency and duration of event-related responses (Friston *et al*. 1998*a*). Latency differences can then be compared with psychophysical differences, such as reaction-time differences. Note, however, that the dispersion derivative is not orthogonal to the canonical form, and although the temporal derivative is orthogonal to the canonical form, this orthogonality will not remain if the model and data are undersampled. Thus, strictly independent tests of latency or duration differences can require explicit orthogonalization of the sampled derivative and dispersion covariates with respect to the sampled canonical form.

(g) *Non-stationary reponses*

The event-related response, or derived effects, can adapt to (interact with) other factors. Whereas the non-linearities described in § 3(b) can be interpreted as interactions between (typically) adjacent event-related responses, non-stationary responses can be regarded as the interaction of other parameters. For example, the response can depend on the task (e.g. the context of a trial), the subject (e.g. the level of arousal), or simply the time during the experiment. Such interactions can be modelled as the product of the parameter and the event-related effect. A particularly interesting case arises when the non-stationarity of a response can be engineered to be orthogonal to stationary components. For example, the variance in a trial of varying duration can be modelled in terms of constant components due to the beginning and end of the trial and a varying component depending on the duration of the trial.

(h) *Random-effect inference*

So far we have assumed that event-related responses are fixed effects; that is, neglecting possible non-stationarity and nonlinearity, the neuronal response from trial to trial can be considered constant. In many cases, however, the actual event-related responses are in fact random effects, i.e. they are themselves drawn from a distribution of possible values. The appropriate null hypothesis is no longer that there is no activation in the experiment but rather that there is no consistent activation in the population of trials from which the present experiment sampled. In this section we consider how including, or failing to include, a random-response component affects the statistical inference drawn from experiments. Much of this discussion reflects the issues raised by Petersson *et al.* (this issue) on multiple-subject inferences (see also Holmes & Friston 1998) (Strange *et al.* 1999).

We consider a hypothetical differential event-related experiment in which the neuronal response varies randomly from trial to trial. We assume that there is in fact no systematic differential activation between trials (i.e. the responses are drawn from the same population distribution with zero mean but non-zero standard deviation), and that the instrument noise is white.

A fixed-effect model for this experiment is the convolution of a delta-function sequence by an HRF. Fitting this model to the data can result in a small but non-zero parameter estimate. Crucially, however, the residual variance is non-white (confounding variability between responses with variability between scans), violating the assumption required to use the residual variance in setting the appropriate significance thresholds. A random-effect model treats each trial separately, in its own partition of the design matrix, meaning that residuals are more likely to be white. To test for the differential effect, the parameter estimates from each trial are taken to a second level of inference, in which their distributions are compared directly. These two stages effectively separate inter-scan and inter-response variability.

An additional advantage of the random-effect model is that, under the null hypothesis, the parameter estimates are likely to be less correlated and hence not require corrections for autocorrelation (§ 5(d)) (although there is no reason why such corrections could not be applied across trials when significant autocorrelation does exist). One problem with separate partitions of the design matrix for each trial is that several degrees of freedom might be required to obtain an accurate estimation of response parameters. This implies several scans per event, with a TR short enough to sample the haemodynamic response sufficiently, and relatively long interstimulus intervals. As shown below, however, long interstimulus intervals are non-optimal for most designs.

6. OPTIMIZATION OF EVENT-RELATED EXPERIMENTAL DESIGNS

In this section we shall discuss the optimization of event-related experimental designs, which is an area of much current interest. Optimization involves maximizing the sensitivity (signal:noise ratio) for particular contrasts (hypotheses) as a function of the stimulus ordering and stimulus onset asynchrony (SOA). For simplicity we assume the noise to be invariant across changes in ordering or SOA, and concentrate on maximizing the signal. We begin by introducing several concepts that are useful in characterizing the space of experimental designs.

(a) *Some definitions*

(i) *Stimulus transition matrices*

An event-related experiment is fully determined by the SOA and a transition matrix. For N different event types, the $N^m \times N$ transition matrix, T, details the probability of a given event type (in the columns of T) conditional on the previous m event types (the rows of T). Table 1 shows a first-order ($m=1$) transition matrix for two randomly ordered event types, A and B (1), a first-order transition matrix for two ordered (alternating) event types (2), and a second-order ($m=2$) transition matrix for two permuted event types (3). Note that the permuted sequence is pseudo-randomized in the sense that it is randomized up to first-order contingencies.

Deterministic designs are those in which the elements of the transition matrix (the probabilities, p_e, of each event type e) are either 0 or 1. Stochastic designs (Heid *et al.* 1997) are those in which $p_e<1$, and a special case of stochastic designs is fully randomized designs in which $p_e=1/N$ for all event types $e=1, \ldots, N$. In most situations, a fully randomized design is desirable (for psychological reasons, for example). Some designs, however, have an inherent non-random ordering (such as alternating transitions between bistable perceptions; see, for example, Kleinschmidt *et al.* 1998). As demonstrated below, different event-type orderings are most sensitive for a given contrast in different ranges of SOA.

Events can be varied in time as well as order. Such designs can be realized by transition matrices in which the rows sum to $p=\Sigma p_e<1$. In the fully randomized design of Dale & Buckner (1997), for example, an additional null trial is introduced (when no event occurs), such that $p_e=1/(N+1)$. The introduction of null trials is equivalent to the stochastic selection of SOAs for the trials of interest (see § 6(c)). Stochastic designs can also be stationary, where p_e is constant over time, or dynamic, where p_e changes over time (K. J. Friston, E. Zarahn,

Table 1. *Examples of (1) random, (2) alternating and (3) permuted (pseudo-random) sequences and their transition matrices*

sequence type		A	B	example
1	A	0.5	0.5	
	B	0.5	0.5	ABBABBBAAABA...
2	A	0	1	
	B	1	0	ABABABABABAB...
3	AA	0	1	
	AB	0.5	0.5	ABBABAABBABA...
	BA	0.5	0.5	
	BB	1	0	

O. Josephs, R. N. A. Henson and A. Dale, unpublished data). An extreme example of a dynamic stochastic design is the blocked-trial activation–rest design, where p_e switches from 0 to 1 in each activation block. We concentrate on deterministic designs below, to illustrate important concepts in the detection of event-related responses, and consider stochastic designs later.

(ii) *Effective HRF*

Although the HRF contains low-frequency components, these components are likely to be contaminated by noise due to subject motion and other long-term drift effects (§3(d)). Its high-frequency components are also redundant when data are smoothed to accommodate temporal autocorrelation in the GLM (§4(e)). It is therefore sensible to introduce the concept of the effective HRF (EHRF). The EHRF is simply the HRF filtered by the band-pass filter comprising the high-pass filter and the temporal smoothing required for statistical inference. By considering this effective response, as opposed to the actual response, we can visualize the measurable component of the haemodynamic response to neuronal activity.

Figure 1*a* shows the EHRF with the use of a canonical HRF (§4(f)), a high-pass filter up to 60 s, and Gaussian smoothing with a full width at half maximum (FWHM) kernel of 4 s. The long side-lobes of the EHRF in the time domain arise from our use of an ideal high-pass filter. Since we are not interested in the time-course of the haemodynamic responses itself, but rather the ability to detect responses that reflect our experimental manipulation, these long time-scale artefacts do not usually concern us. However, this does provide a mechanism whereby, if the model for the event-related response is incomplete, power evoked by one event can 'leak' into a region of time that is modelled by another event (an issue that needs further exploration). The effect of temporal smoothing can be seen in the attenuation of the high-frequency components in the frequency domain. The optimization of an experimental design can then be viewed as the maximization of experimental power under the frequency profile of the EHRF.

(iii) *Inter-effect interval*

For deterministic designs, it is the inter-effect interval (IEI) that determines optimal sensitivity for a particular contrast. The IEI is the time between repetitions of event types with the same contrast weight. The IEI for a simple difference between two alternating event types (a [1 −1] contrast), for example, is twice the SOA, as is the IEI for the interaction between four alternating event types in a 2×2 factorial design (a [1 −1 −1 1] contrast), whereas the IEI for the main effect (a [1 1] or [1 1 1 1] contrast, respectively) is equal to the SOA. As shown below, the optimal SOA for a given contrast in a deterministic design is that with an experimental frequency, 1/IEI, that is closest to the peak of the EHRF power spectrum.

(iv) *Estimated measurable power*

The signal magnitude associated with a particular contrast and a particular experimental design can be indexed by the estimated measurable power (EMP). The EMP can be simulated by taking the linear neuronal model (§4(c)), multiplying by the contrast of interest, convolving with the EHRF (again assuming linearity of the haemodynamic response; §4(f)(ii)), sampling every scan (TR) and calculating the total energy (sum of squared signal across scans). The EMP is the total energy divided by the number of scans (to normalize for different experimental durations). Assuming that the noise level is fixed (i.e. not influenced by the parameters being varied), then the experimental design with the highest EMP is likely to be the most sensitive for that contrast. EMP is therefore proportional to the efficiency of a design for a particular contrast, which is proportional to the variance of a column in a design matrix (K. J. Friston, E. Zarahn, O. Josephs, R. N. A. Henson and A. Dale, unpublished data). For multiple contrasts, or contrasts involving multiple basis functions, the efficiency of a design can be characterized by the trace of the covariance of the contrasted design matrix. The EMPs for single contrasts in some example designs are discussed below.

(b) *Simulated optimization*

Figure 2 shows EMP as a function of SOA for the main and differential effects in various orderings of two event types. These simulations involved 10 000 events (to ensure stable estimates for randomized designs), the EHRF shown in figure 1*a*, and continuous sampling with a TR of 1 s. We emphasize that resulting EMP is not necessarily linearly related to sensitivity (for example the resulting *t*-value)—we assume only that EMP is a monotonic index of sensitivity—and that for SOAs of less than 2 s (Friston *et al*. 1998*b*), nonlinear (saturation) terms will prevent the EMP from growing without limit.

Figure 2*a* shows sensitivity for the random, alternating and permuted designs in table 1. Maximal sensitivity to a differential effect (a [1 −1] contrast) arises with randomized designs and short SOAs. Maximal sensitivity to a main effect (a [1 1] contrast) is of course independent of stimulus ordering, and occurs with an SOA of *ca.* 18 s (i.e. with a dominant experimental frequency close to that of the EHRF). Thus we immediately see that whether one's hypothesis concerns a differential effect or a main effect determines whether one should use a short or a long SOA respectively (in deterministic designs).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

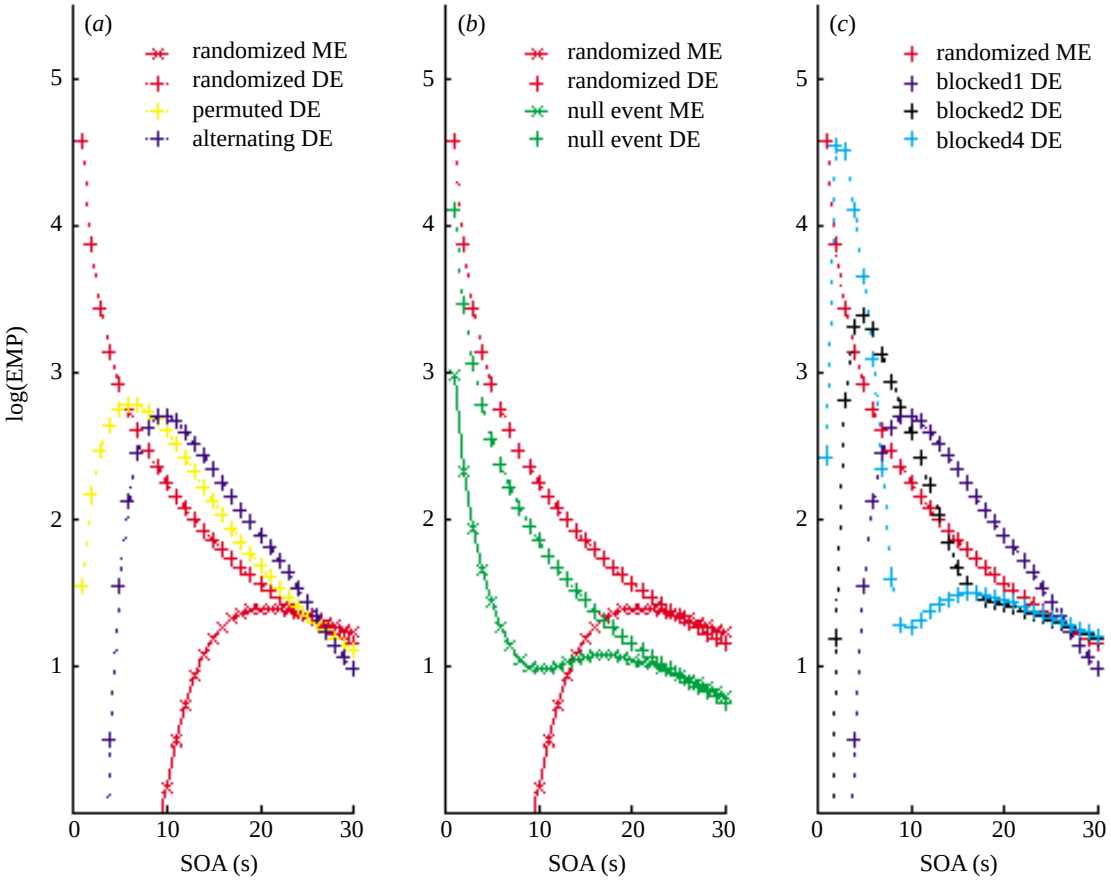

Figure 2. Logarithm of the scaled EMP against SOA for the main effect (main effect, ME, solid lines) and differential effect (differential effect, DE, broken lines) for two event types in (*a*) randomized, alternating and permuted designs (table 1), (*b*) randomized designs with and without null events (§6(b)), and (*c*) randomized designs and blocked designs of one (alternating), two or four events.

At short SOAs, the main effect of two event types (the constant component) is effectively a raised baseline that is removed by the high-pass filter. For alternating designs, the high-frequency components of the differential effect at short SOAs are lost by the temporal smoothing inherent in the HRF and smoothing kernel. For randomized designs, the differential effect is actually enhanced at short SOAs through the summation of overlapping HRFs (see, for example, Burock *et al*. 1998). In fact, the maximum energy for such rapid, randomized designs actually occurs in the scans after runs of the same event type (see figure 1*b* and §6(d)).

At intermediate SOAs between 6 and 25 s, the optimal design for a differential effect is an alternating or permuted design. With such designs, the optimal SOA is *ca*. 9 s, giving an IEI (twice the SOA; §6(a)(iii)) that is close to the dominant components of the EHRF. More generally, the optimal SOA for a given contrast in an alternating design is that with an IEI close to 18 s (with the present parametrization of high-pass and low-pass filtering). Thus the optimal SOA for a pairwise difference between two of four alternating event types (a simple effect, or a [1 −1 0 0] contrast) would be *ca*. 4.5 s. The decreased sensitivity of randomized designs in these intermediate SOA regimes arises because low-frequency components resulting from long runs of an event type are attenuated by the high-pass filter. An alternating sequence might not, of course, be appropriate for psychological considerations, so a permuted design, which has a similar sensitivity profile but is randomized to first order, might be a good compromise.

Figure 2*b* shows the sensitivity for a design suggested by Dale & Buckner (1997). This fully randomized design involves a third null event, in which no stimulus occurs (i.e. the two event types are effectively randomized in time as well as order). If null events are treated as a third event type, main and differential effects can be simulated by [1 1 0] and [1 −1 0] contrasts, respectively. This design affords improved sensitivity to the main effect even at short SOAs. Thus, although this design is non-optimal for differential contrasts, relative to a randomized design with only two event types, it is a good design for retaining sensitivity to both main and differential effects at short SOAs.

Figure 2*c* shows the sensitivity of short, blocked designs, with alternating blocks of one, two or four events of each type. These blocked designs are more sensitive than the randomized design at short SOAs, particularly as block length increases (at least until the frequency of block alternation approaches the high-pass cut-off (Hutton *et al*. 1998)). Again, this is because, under the linearity assumption, responses to events of the same type are summed, concentrating power at the block alternation frequency.

(c) *Analytical optimization*

Many of the simulated results in the previous section can be confirmed by analytical methods. Friston *et al.* (K. J. Friston, E. Zarahn, O. Josephs, R. N. A. Henson and A. Dale, unpublished data), for example, characterized the sensitivity to responses of a single event type as a function of the minimal SOA, SOA_{min}, and the probability, p, of an event occurrence every SOA_{min}. Under the linearity assumption, optimal sensitivity occurs in a stationary stochastic design with $p = 0.5$ and minimal SOA_{min}. Thus, contrary to the deterministic cases considered above, long SOAs are not necessary to measure main effects if stochastic designs are employed: the random 'jitter' in time introduces lower-frequency components that survive smoothing, much like the Dale–Buckner null events in figure 2*b*. Future investigation of event-related optimization will need to parameterize and quantify the influence of nonlinear components of neuronal and haemodynamic responses on sensitivity, particularly for short SOAs and for variable durations of events.

(d) *Interpretation in rapid, randomized designs*

Figure 1*b* shows a section of the simulated signal in a differential contrast with a randomized design and an SOA of 2 s. The bottom plot, showing the instantaneous power at each scan point, illustrates that, at short SOAs, runs of events of the same type generate large amounts of energy. Conversely, periods of alternating event types produce minimal energy. This observation has implications for the interpretation of any significant differential effect: any such effect is driven mainly by a difference between responses to event type A and event type B in the context of runs of A or B. In other words, we are left with the same type of context-dependent interpretation that is the lacuna of blocked designs. For example, if a negative priming effect existed for the immediate repetition of event type A but not B (such that the response to A diminished with immediate repetition), the finding of a significant difference between event types A and B might apply only to primed events, and there might not in fact be any difference between unprimed event types A and B. This is one reason why the greater EMP for short SOAs might be sacrificed in favour of the clearer interpretation afforded by longer SOAs in which haemodynamic overlap is reduced.

7. EXAMPLE EVENT-RELATED EXPERIMENTAL DESIGNS

Now that we have covered measurement, modelling, inference and optimization, we can discuss how these techniques have been used to address neuroscientific questions.

(a) *Single event type*

Josephs *et al.* (1997) used a Fourier basis set of 16 components up to 0.5 Hz (§ 4(g)) to model the response to words heard every 33 s. An *F*-test revealed significant activation in auditory and periauditory regions. Plots of the event-related response peaked at 5–8 s after the stimulus, followed by an undershoot 10 and 25 s after the stimulus. Zarahn *et al.* (1997*a*) used a set of time-shifted, empirically derived HRFs to model responses during experimental and control trials of a visuospatial working memory task presented every 30 s. Paired *t*-tests on individual HRFs revealed significant activations in sensorimotor and prefrontal regions associated with temporally distinct perceptual, mnemonic and response components of each experimental trial. Dale & Buckner (1997) used gamma functions (parameterized by Boynton *et al.* (1996)) to model the response to 1 s bursts of one, two or three flickering visual chessboards. Significant activation was observed in the primary visual cortex, with responses within bursts summing linearly even for SOAs as short as 2 s. These examples illustrate the range of different modelling techniques that have been used to detect significant event-related responses.

(b) *Raised baseline*

The two principal disadvantages of multiple-trial experiments with a single event type relate to the predictability of events and the lack of any high-level control. The subject knows in advance what type of event they are about to experience and a voxel might be activated not because of the effect of interest (e.g. a word being presented) but because the event is more salient, or attention-grabbing, than the inter-trial baseline. One might even detect event-related deactivations in brain regions that reflect, for example, a growing anticipation of an event. A raised-baseline technique might suffice in some cases (§ 5(e)). This form of experiment has been used (R. Vandenberghe, O. Josephs, L. K. Tyler, C. J. Price, R. Turner, R. S. J. Frackowiak and K. J. Friston, unpublished data) when the event of interest was temporally extended, comprising a short sequence of seven words (lasting 3.85 s) within a background sequence of consonant letter strings presented every 550 ms. The prolonged HRF was modelled with a Fourier basis set (§ 4(g)). As expected, responses associated with word-processing were identified in the left supramarginal, superior temporal and Sylvian fissure regions (although these responses did not seem to differentiate between meaningful sentences and random word sequences).

A similar design has been used (B. A. Strange, R. N. A. Henson, K. J. Friston and R. J. Dolan, unpublished data) in which sequences of semantically related neutral-context words were presented at a rate of one word every 3 s (to allow subjects to make a judgement about each word). Three 'oddball' words were randomly inserted into these sequences: a perceptual 'oddball' (a word presented in a different font), a semantic 'oddball' (a semantically unrelated word) and an emotional 'oddball' (a word with aversive emotional connotations). The slower presentation rate in this case meant that a tonic baseline might not have been reached, so context words were explicitly modelled as confounds (§ 5(e)). Enhanced responses relative to context words were identified in the fusiform cortex for perceptual 'oddballs', in the left prefrontal cortex for semantic 'oddballs', and in the amygdala for emotional 'oddballs'.

(c) *Simple differential response*

Most event-related experiments use two or more event types and are primarily interested in the differential effect between event types (§ 5(f)(i)). A popular design for differential effects is the rapid, randomized

presentation with null events developed by Dale & Buckner (1997). Buckner *et al*. (1998), for example, used this design to examine object priming by comparing responses to novel object pictures with responses to previously studied pictures. Several brain regions, including the extrastriate visual, inferior temporal and left dorsal prefrontal cortices, showed reduced responses to familiarized compared with novel pictures. One advantage of this randomization scheme is that plots of the differential response between each event type and fixation (e.g. novel pictures compared with null events) can be generated simply by selective averaging of the data (although statistical tests of these differences remain best achieved with temporal basis functions).

Wagner *et al*. (1998) reported another use of this design to examine differential responses to words presented visually as a function of whether or not the words were subsequently remembered. Enhanced responses in the left posterior prefrontal, parahippocampal and fusiform cortices predicted which words were later recognized. This experiment illustrates one of the important advantages of event-related designs (§2(a)), namely the ability to classify event types post hoc as a function of the subject's behaviour. However, one associated problem is to ensure that the order of the subjectively classified event types remains random (for example, the well-established tendency for stimuli towards the start and end of study lists to be better remembered might mean that remembered events tend to co-occur early and late in the scanning period).

Another example of a subjective classification of events was reported by Henson *et al*. (1999*a*), in which subjects indicated the nature of their conscious experience when trying to remember words studied previously. Words that produced a clear recollection of their study episode were given a 'remember' (R) judgement, whereas words that produced a feeling of familiarity in the absence of recollection were given a 'know' (K) judgement. R judgements were associated with enhanced responses in left anterior prefrontal and superior parietal cortices, whereas K judgements were associated with enhanced responses in right dorsal prefrontal and bilateral anterior cingulate cortices. This study also effected a univariate random-effects inference across subjects (§5(h)) by reducing event-related responses to a single parameter (the height of a canonical HRF). A final example of a differential effect that can be indexed only by subjectively defined events is illustrated by Kleinschmidt *et al*. (1998), who scanned subjects while they viewed perceptually bistable figures. By comparing button presses made by subjects when a percept was stable with button presses made when the percept spontaneously reversed, these authors identified increased responses associated with reversals in the ventral occipital and intraparietal cortex, and decreased responses in the primary visual cortex and the pulvinar, even though the visual stimulus remained constant throughout. Event-related experiments such as these are therefore beginning to identify the neural correlates of different types of conscious experience.

(d) *Factorial responses*

Fewer studies have used factorial, event-related designs (§5(f)(ii)). One example is a study (R. N. A. Henson, T. Shallice and R. J. Dolan, unpublished data) that examined the interaction between repetition priming within an experimental context (as in Buckner *et al*. (1998), discussed above) and whether or not the stimulus was familiar before the experiment. In one experiment, the interaction was between first and second presentations of famous and non-famous (unfamiliar) faces. A region in the right lateral fusiform cortex showed a decreased response to the repetition of famous faces, comparable to the results of Buckner *et al*. (1998), but an increased response to the repetition of non-famous faces. Furthermore, the same region showed the same interaction in a second experiment in which the stimuli were simple line-drawings that were either familiar (symbols) or unfamiliar (rearranged symbols). One possibility is that the enhanced response to the repetition of unfamiliar stimuli in this region reflects the formation of new stimulus representations.

(e) *Parametric responses*

In the study of repetition priming described above (R. N. A. Henson, T. Shallice and R. J. Dolan, unpublished data), the first and second presentations of familiar and unfamiliar stimuli were randomly intermixed throughout the scanning period. This entailed a range of lags (number of intervening stimuli) between the first and second occurrence of a specific stimulus. The parametric effect of this repetition lag was examined by multiplying the event-related covariate for the second occurrence of stimuli by its mean-corrected lag (§5(f)(iii)). Although the main effect of lag is confounded by time (because larger lags are necessarily associated with stimuli occurring towards the start and the end of the scanning session), the interaction between lag and stimulus familiarity is not (illustrating the value of factorial designs; §5(f)(ii)). The same region that showed an interaction between face repetition and face familiarity also showed a modulation by lag, such that the response to the second presentation of famous faces increased with lag, whereas the response to the second presentation of a non-famous face decreased with lag. This interaction suggests that both the repetition inhibition for famous faces and the repetition facilitation for non-famous faces were short-lived.

(f) *Response parameterization*

The use of multiple basis functions can permit tests of different aspects of the haemodynamic response. The use of temporally shifted gamma functions, for example, allowed Schacter *et al*. (1997) to differentiate between early- and late-onset HRFs, with the best-fitting function for the anterior prefrontal cortex being delayed relative to that for the visual cortex during a visual recognition task (although this delay cannot be unequivocally attributed to haemodynamic or neural differences; §3(a)(ii)). Less equivocal latency differences in neuronal activity can be inferred from differences between fitted responses for event types within the same brain region. Friston *et al*. (1998*a*), for example, illustrated how the use of a canonical HRF and its first-order derivative with respect to time allows one to test for differences in response latency. The use of the derivative of a canonical HRF with respect to its dispersion similarly allows one to test for differences in response duration.

The temporal derivative of a canonical HRF has been used (Henson *et al*. 1999*c*) to test for latency differences in a lexical decision task. Regions that exhibited enhanced responses to words relative to non-words (greater height parameter estimates), including the left angular and left inferior temporal cortex, also exhibited shorter latencies (greater derivative parameter estimates). This could reflect nonlinearities in the underlying haemodynamic response (Vasquez & Noll 1998), such as the non-independence of response magnitude and response onset, or it could reflect a true difference in the onset of neural activity after words and non-words.

Another advantage of using the derivative to test for latency differences is that it provides an approximation to the size of the latency difference. By comparison with the first-order Taylor expansion of a haemodynamic function, $f(t)$,

$$f(t+\mathrm{d}t) \approx f(t) + f'(t)\mathrm{d}t,$$

where f' is the derivative of f with respect to time, the parameter estimates b_1 and b_2 obtained from fitting a canonical basis function $g_1(t)$ and its temporal derivative $g_2(t)$ (§5(f)(iv)),

$$f(t) \approx b_1 g_1(t) + b_2 g_2(t),$$

allow an approximation of $\mathrm{d}t$ as

$$\mathrm{d}t \approx b_2/b_1,$$

(assuming the canonical HRF fits well, i.e. $b_1 > 0$). With the use of this approximation, values of $\mathrm{d}t$ have been found (Henson *et al*. 1999*c*) that differed by *ca*. 0.3–3 s in brain regions that were differentially responsive to words and non-words. Latencies of this size are likely to reflect haemodynamic rather than neural differences (given that reaction times were less than 1 s for both word and non-word decisions). This use of multiple basis functions is an alternative to the nonlinear fitting of parametrized functions, such as the delay and dispersion of a Gaussian HRF (Krueggel & Von Cramon 1998).

(g) *Non-stationary responses*

An example of the analysis of non-stationary responses is provided by Buechel *et al*. (1998), who sought to characterize changes in event-related responses during aversive conditioning of a neutral face (CS+) by a loud noise (US). By using a partial reinforcement schedule, the main events of interest were presentations of the CS+ in the absence of the US (unpaired CS+ events). Covariates sensitive to the learning of the CS–US contingency were created by convolving the unpaired CS+ event trains with a canonical HRF and multiplying by an exponentially decreasing function of time. The only region showing such a response–time interaction was the left amygdala, suggesting that this structure shows adaptation to the CS+ during conditioning.

This study also illustrates another advantage of event-related designs: in previous conditioning studies, blocks of CS+ stimuli have been compared with blocks of CS− stimuli (stimuli that are never followed by the US). To avoid the confound of the US itself, however, these studies have needed to present CS+ stimuli alone during the scanning window. Unfortunately, this means that CS+ responses are measured in the context of an extinction schedule, a confound explicitly acknowledged by such studies (for example, Morris *et al*. 1998). Event-related designs with a partial reinforcement schedule, like that of Buechel *et al*. (1998), permit the separation of intermixed trials of CS+ with the US, and CS+ without the US.

(h) *Responses as random effects*

The modelling of event-related responses as random effects is a relatively recent suggestion and is an important consideration if the variability of effect size (e.g. response magnitude) is large relative to the interscan variability. This type of analysis has recently been applied (Josephs & Friston 1999) to epileptic spikes (triggered between scans in a 'burst-mode' procedure (Josephs *et al*. 1999)), which is one situation in which a large inter-response variability would seem likely. Several brain regions showed a significant difference in the mean parameter estimate for spike events and the mean parameter estimate for control events. We are currently investigating the relative size of inter-response and interscan variability, to address design issues concerning the appropriate number of events and scans for effective statistical power.

8. CONCLUSIONS

Having summarized the important advantages of event-related designs and the measurement of event-related haemodynamic responses with the use of echo-planar fMRI, we have discussed current issues in the modelling, inference and optimization of event-related designs. Future development in the analysis of efMRI would seem to require further investigation of (i) the reproducibility of the haemodynamic response, especially over different brain regions and different subjects (§3(a)); (ii) the degree of nonlinearity in the measured response, which has particular relevance to the appropriate modelling of experimental variance (§4) and the optimization of event-related designs with short SOAs (§6(c)); and (iii) the appropriateness of modelling event-related responses as random, rather than fixed, effects (§5(h)), which has implications for the statistical power of experimental designs.

This work was supported by the Wellcome Trust. We thank Karl Friston for useful comments on an earlier draft.

REFERENCES

Aguirre, G. K., Zarahn, E. & D'Esposito, M. 1997 Empirical analyses of BOLD fMRI statistics. II. Spatially smoothed data collected under null-hypothesis and experimental conditions. *NeuroImage* **5**, 199–212.

Aguirre, G. K., Zarahn, E. & D'Esposito, M. 1998 The variability of human, BOLD hemodynamic responses. *NeuroImage* **8**, 360–369.

Boynton, G. M., Engel, S. A., Glover, G. H. & Heeger, D. J. 1996 Linear systems analysis of functional magnetic resonance imaging data in human V1. *J. Neurosci*. **16**, 4207–4221.

Buckner, R. L., Goodman, J., Burock, M., Rotte, M., Koustaal, W., Schacter, D., Rosen, B. & Dale, A. M. 1998 Functional–anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. *Neuron* **20**, 285–296.

Buechel, C., Wise, R. J. S., Mummery, C. J., Poline, J.-B. & Friston, K. J. 1996 Nonlinear regression in parametric activation studies. *NeuroImage* **4**, 60–66.

Buechel, C., Morris, J., Dolan, R. J. & Friston, K. J. 1998 Brain systems mediating aversive conditioning: an event-related fMRI study. *Neuron* **20**, 947–957.

Burock, M. A., Buckner, R. L., Woldorff, M. G., Rosen, B. R. & Dale, A. M. 1998 Randomised event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *NeuroReport* **9**, 3735–3739.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J. & Smith, E. E. 1997 Temporal dynamics of brain activation during a working memory task. *Nature* **386**, 604–607.

Corfield, D. R., Murphy, K., Josephs, O., Adams, L. & Turner, R. 1998 Modulation by hypercapnia of activation-related BOLD signal. *NeuroImage* **7**, 259.

Dale, A. M. & Buckner, R. L. 1997 Selective averaging of rapidly presented individual trials using fMRI. *Hum. Brain Mapp.* **5**, 329–340.

D'Esposito, M. D., Zarahn, E. & Aguirre, G. K. 1999 Event-related functional MRI: implications for cognitive psychology. *Psychol. Bull.* (In the press.)

DeYoe, E. A., Neitz, J., Bandettini, P. A., Wong, E. C. & Hyde, J. S. 1992 Time-course of event-related MR signal enhancement in visual and motor cortex. In *Proceedings, International Society of Magnetic Resonance in Medicine, Inc. 11th Annual Meeting, Berlin*, p. 1824.

Friston, K. J., Frith, C. D., Turner, R. & Frackowiak, R. S. J. 1995*a* Characterizing evoked hemodynamics with fMRI. *NeuroImage* **2**, 157–165.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. B., Frith, C. D. & Frackowiak, R. S. J. 1995*b* Statistical parametric maps in functional imaging; a general linear approach. *Hum. Brain Mapp.* **2**, 189–210.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D. & Turner, R. 1998*a* Event-related fMRI: characterising differential responses. *NeuroImage* **7**, 30–40.

Friston, K. J., Josephs, O., Rees, G. & Turner, R. 1998*b* Nonlinear event-related responses in fMRI. *Magn. Reson. Med.* **39**, 41–52.

Heid, O., Gonner, F. & Schroth, G. 1997 Stochastic functional MRI. *NeuroImage* **5**, 476.

Henson, R. N. A., Rugg, M. D., Shallice, T., Josephs, O. & Dolan, R. J. 1999*a* Recollection and familiarity in recognition memory: an event-related fMRI study. *J. Neurosci.* **19**, 3962–3972.

Henson, R. N. A., Shallice, T., Price, C., Dolan, R. J., Friston, K. & Turner, R. 1999*b* Lexical decision: differences in magnitude and onset as indexed by event-related fMRI. *NeuroImage* **9**, 1044.

Henson, R. N. A., Buechel, C., Josephs, O. & Friston, K. 1999*c* The slice-timing problem in event-related fMRI. *NeuroImage* **9**, 125.

Holmes, A. P. & Friston, K. J. 1998 Generalisability, random effects and population inference. *NeuroImage* **7**, 754.

Holmes, A. P., Josephs, O., Buechel, C. & Friston, K. J. 1997 Statistical modelling of low frequency confounds in fMRI. *NeuroImage* **5**, 480.

Hutton, C., Friston, K. & Turner, R. 1998 The effect of interstimulus interval on signal response in fMRI. *NeuroImage* **7**, 591.

Johnson, M. K., Nolde, S. F., Mather, M., Kounios, J., Schacter, D. L. & Curran, T. 1997 The similarity of brain activity associated with true and false recognition depends on test format. *Psychol. Sci.* **8**, 250–257.

Josephs, O. & Friston, K. 1999 Event-related fMRI: which error variance measure for statistical inference? *NeuroImage* **9**, 186.

Josephs, O., Turner, R. & Friston, K. 1997 Event-related fMRI. *Hum. Brain Mapp.* **5**, 243–248.

Josephs, O., Lemieux, L., Krakow, K. & Friston, K. 1999 Burst mode event-related fMRI. *NeuroImage* **9**, 219.

Kleinschmidt, A., Buechel, C., Zeki, S. & Frackowiak, R. S. J. 1998 Human brain activity during spontaneously reversing perception of ambiguous figures. *Proc. R. Soc. Lond.* B **265**, 2427–2433.

Kruggel, F. & von Cramon, D. Y. 1998 Modeling the hemodynamic response in single trial fMRI experiments. Technical report 12/98. Leipzig: Max-Planck-Institute of Cognitive Neuroscience.

Lee, A. T., Glover, G. H. & Meyer, C. H. 1995 Discrimination of large venous vessels in time-course spiral blood-oxygenation-level-dependent magnetic-resonance functional neuroimaging. *Magn. Reson. Imag. Med.* **33**, 745–754.

McCarthy, G., Puce, A., Constable, R. T., Krystal, J. H., Gore, J. C. & Goldman-Rakic, P. 1996 Activation of human prefrontal cortex during spatial and nonspatial working memory tasks measured by functional MRI. *Cerebr. Cortex* **6**, 600–611.

McGonigle, D. J., Howseman, A. M., Athwal, B. S., Friston, K. J., Frackowiak, R. S. J. & Holmes, A. P. 1999 An examination of inter-session differences in fMRI. *NeuroImage* **9**, 38.

Morris, J. S., Ohman, A. & Dolan, R. J. 1998 Conscious and unconscious emotional learning in the human amygdala. *Nature* **393**, 467–470.

Pollmann, S., Wiggins, C. J., Norris, D. G., Von Cramon, D. Y. & Schubert, T. 1998 Use of short intertrial intervals in single-trial experiments: a 3T fMRI study. *NeuroImage* **8**, 327–339.

Price, C. J., Veltman, D. J., Ashburner, J., Josephs, O. & Friston, K. J. 1999 The critical relationship between the timing of stimulus presentation and data acquisition in blocked designs with fMRI. *NeuroImage*. (In the press.)

Richter, W., Andersen, P., Georgopoulos, A. P. & Kim, S. G. 1997 Time-resolved fMRI measurement of brain activity during a delayed motor task. In *Proceedings, International Society of Magnetic Resonance in Medicine, Inc. 5th Scientific Meeting, Vancouver*, p. 11.

Rosen, B. R., Buckner, R. L. & Dale, A. M. 1998 Event-related functional MRI: past, present and future. *Proc. Natl Acad. Sci. USA* **95**, 773–780.

Schacter, D. L., Buckner, R. L., Koustaal, W., Dale, A. M. & Rosen, B. R. 1997 Late onset of anterior prefrontal activity during true and false recognition: an event-related study. *NeuroImage* **6**, 259–269.

Strange, B. A., Portas, C. M., Dolan, R. J., Holmes, A. P. & Friston, K. J. 1999*b* Random effects analyses for event-related fMRI. *NeuroImage* **9**, 36.

Vasquez, A. L. & Noll, D. C. 1998 Nonlinear aspects of the BOLD response in functional MRI. *NeuroImage* **7**, 108–118.

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., Rosen, B. R. & Buckner, R. L. 1998 Building memories: remembering and forgetting of verbal experience as predicted by brain activity. *Science* **281**, 1188–1191.

Worsley, K. J. & Friston, K. J. 1995 Analysis of fMRI time-series revisited—again. *NeuroImage* **2**, 173–181.

Zarahn, E., Aguirre, G. K. & D'Esposito, M. 1997*a* A trial-based experimental design for fMRI. *NeuroImage* **6**, 122–138.

Zarahn, E., Aguirre, G. K. & D'Esposito, M. 1997*b* Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage* **6**, 179–197.