CHAPTER

15

Efficient Experimental Design for fMRI

R. Henson

INTRODUCTION

This chapter begins with an overview of the various types of experimental design, before proceeding to various modelling choices, such as the use of events versus epochs. It then covers some practical issues concerning the effective temporal sampling of blood oxygenation-level-dependent (BOLD) responses and the problem of different slice acquisition times. The final and main part of the chapter concerns the statistical efficiency of functional magnetic resonance imaging (fMRI) designs, as a function of stimulus onset asynchrony (SOA) and the ordering of different stimulus-types. These considerations allow researchers to optimize the efficiency of their fMRI designs.

TAXONOMY OF EXPERIMENTAL DESIGN

Most experiments involve the manipulation of a number of factors over a number of levels. For example, a factor of spatial attention might have two levels of left versus right covert attention (relative to fixation), while a second factor might be whether a stimulus is presented to the left or right visual hemi-field. Orthogonal manipulation of these two factors corresponds to a '2×2' 'factorial' design, in which each factor-level combination constitutes an experimental condition (i.e. four conditions in this case; see Chapter 13). Factors with a discrete number of levels, as in the above example, are often called 'categorical'. Other factors may have continuous values (such as the duration of the stimulus for example), and may have as many 'levels' as there are values. Such factors are called 'parametric'. Below, we discuss briefly different designs in the context of the general linear model (GLM) and some of the assumptions they entail.

Single-factor subtraction designs and 'pure insertion'

The easiest way to illustrate different types of design is with examples. Plate 15(a) (see colour plate sections) shows an example design matrix with 12 conditions and 5 sessions (e.g. 5 subjects). The data could come from a positron emission tomography (PET) experiment or from a second-level analysis of contrast images from an fMRI experiment. We use this example to illustrate a number of designs and contrasts below. Initially, we will assume that there was only one factor of interest, with two levels (that happened to occur six times in alternation). These might be reading a visually-presented cue word ('Read' condition) and generating a semantic associate of the cue ('Generate' condition). If one were interested in the brain regions involved in semantic association, then one might subtract the Read condition from the Generate condition, as shown by the *t*-contrast in Plate 15(a). The logic behind this subtraction is that brain regions involved in processes common to both conditions (such as visual processing of the cue word) will be equally active in both conditions, and therefore not appear in the resulting statistical parametric mapping (SPM). In other words, the contrast should reveal activations related to those processes unique to generating semantic associations, relative to reading words.

A criticism often levelled at such 'cognitive subtractions' is that the conditions may differ in ways other than those assumed by the specific cognitive theory under investigation. For example, the Generate and Read conditions might differ in phonological processes, as well as semantic processes (i.e. the subtraction is 'confounded'). The assumption that tasks can be elaborated so that they call upon a single extra process is called the 'pure insertion' assumption, and has been the source of much debate in neuroimaging (Friston *et al.*, 1996). In fact, the debate goes back to the early days of experimental psychology, e.g. the 'Donders' method of subtraction and its

Statistical Parametric Mapping, by Karl Friston et al.
ISBN–13: 978-0-12-372560-8 ISBN–10: 0-12-372560-7

subsequent refinements (Sternberg, 1969). In short, the concerns about the pure insertion assumption are not unique to neuroimaging (Henson, 2005). Below we will consider some ways to ameliorate such concerns.

Cognitive conjunctions

One way to minimize the probability that interpretation of activation is confounded is to isolate the process of interest using multiple different subtractions. The probability of each subtraction being confounded by the same (uninteresting) differences is thus reduced. In other words, one only considers activation that is common to all subtractions: a method called 'cognitive conjunction' (Price and Friston, 1997). For example, consider an experiment with four conditions (Plate 16): passively viewing a colour-field (Viewing Colour), naming the colour of that field (Naming Colour), passively viewing an object (Viewing Object), and naming an object (Naming Object). One might try to isolate the neuronal correlates of visual object recognition by performing a conjunction of the two subtractions: (1) Object versus Colour Viewing and (2) Object versus Colour Naming. Both subtractions share a difference in the stimulus (the presence or absence of an object), but differ in the nature of the tasks (or 'contexts'). Thus a potential confound, such as number of possible names, which might confound the second subtraction, would not necessarily apply to the first subtraction, and thus would not apply to the conjunction as a whole.

The precise (statistical) definition of a conjunction has changed with the history of SPM, and different definitions may be appropriate for different contexts (the details are beyond the present remit, but for further discussion, see Friston *et al.*, 2005; Nichols *et al.*, 2005). In the present context of 'cognitive' conjunctions, a sufficient definition is that a region survives a statistical threshold in all component subtractions ('inclusive' masking), with the possible further constraint of no interaction between the subtractions ('exclusive' masking). A posterior temporal region shows this type of pattern in Plate 16 (upper panel) and might be associated with implicit object recognition.

A single parametric factor

To illustrate a parametric factor, let us return to the Generate and Read experiment in Plate 15. One might be interested whether there is any effect of time during the experiment (e.g. activation may decrease over the experiment as subjects acquire more practice). In this case, a time factor can be modelled with 12 discrete levels, over which the effects of time could be expressed in a number of different ways. For example, time may have a linear effect, or it may have a greater effect towards the start than towards the end of the experiment (e.g. an exponential effect). The *t*-contrast, testing the former linear effect – more specifically, for regions showing a decrease in activation over time – is shown in Plate 15(b) (in fact, the plot of activity in the highlighted region suggests an exponential decrease, but with a sufficiently linear component that it is identified with the linear contrast).

When the precise function relating a parametric experimental factor to neuronal activity is unknown, one option is to express the function in terms of a polynomial expansion, i.e.:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots \quad \textbf{15.1}$$

where β_i are the parameters to be estimated. For N levels of a factor, the expansion is complete when the terms run from *0*th-order up to order $N-1$. In the latter case, the corresponding design matrix is simply a rotation of a design matrix where each level is modelled as a separate column. An example design matrix for an expansion up to second-order, over 12 images, is shown in Plate 17(a) (e.g. for a single subject in Plate 15): the first column models linear effects, the second column models quadratic effects, and the third column models the 0th-order (constant) effect. An *F*-contrast on the second column identifies a region that shows an inverted-U shape when activity is plotted as a function of the 12 levels of the factor. If this factor were rate of word generation, for example, one might conclude that activity in this region increases as the word rate increases to a certain level, but then decreases if that (optimal) level is surpassed. Parametric modulations that have only one level per value (i.e. are modelled as continuous rather than discrete values) can be modelled by a 'parametric modulation' in SPM. An example of a parametric modulation of event-related responses is shown in Chapter 14.

Factorial designs

Many experiments manipulate more than one factor concurrently. When each condition is associated with one level of every factor, it is called a 'factorial' design. These are common in experimental sciences because they allow tests of not only differences between the levels of each factor, collapsing over other factors ('main effects'), but also how the effect of one factor depends on the level of another factor ('interactions'). Let us return to the object–colour experiment in Plate 16. This experiment can conceived as a '2×2' design, where one factor, Task, has two levels (viewing versus naming) and the other, Stimulus, also has two levels (colour-field or

object). This '2-way' design, therefore, offers tests of two main effects and one interaction (see Chapter 13 for a generalization to '*M*-way' factorial designs). The component subtractions considered for the cognitive conjunction above are sometimes called the 'simple effects'. The interaction in this design would test where the difference between objects and colour-fields varies between a naming task and a viewing task. If these conditions were ordered: Viewing Object, Viewing Colour, Naming Object, Naming Colour (i.e. with the Task factor 'rotating' slowest), then the interaction would have contrast weights[1 −1 −1 1]. This can be conceived as the difference of two differences, i.e. difference of the two simple effects, i.e. [1 −1 0 0]−[0 0 1 −1], or as the 'product' of two differences, i.e. [1 −1]⊗[1 −1], where ⊗ is the Kronecker product.

When testing one tail of the interaction (i.e. with a *t*- rather than *F*-contrast), namely where objects produce greater activation relative to colour-fields when named, rather than when viewed, a region was found in temporal cortex (see Plate 16 – lower SPM), anterior to that in the conjunction (upper SPM). Given that the region showed little difference between objects and colour-fields under passive viewing (i.e. this simple effect was not significant), the pattern in Plate 16 might be termed 'naming-specific object-recognition'. Note also that, if one attempted to isolate visual object-recognition using only a naming task, this interaction could be used as evidence of a failure of pure insertion, i.e. that naming an object in the visual field involves more than simply visual recognition (Price and Friston, 1997).

An example of an interaction involving a parametric factor is shown in Plate 17(b). This contrast tests for a linear time-by-condition interaction in the Generate-Read experiment (when conceived as a 2 × 6 factorial design). Again, the contrast weights can be viewed as the Kronecker product of the Generate versus Read effect and the linear time effect, i.e. [1 −1]⊗[5 3 1 −1 −3 −5]. This *t*-contrast asks where in the brain the process of semantic association decreases (linearly) over time (as might happen, for example, if subjects showed stronger practice effects on the generation task than the read task).

A final example of an interaction is shown in Figure 15.1. In this case, the effects Task (Generate versus Read), Time, and their interactions have been expressed in the design matrix (for a single subject), rather than in the contrast weights (cf. Plate 17(a)). This illustrates the general point that one can always re-represent contrasts by rotating both the design matrix and the contrast weights (see Chapter 13 for further discussion). More precisely, the columns of the design matrix in Figure 15.1 model (from left to right): effect of Task, linear then quadratic effects of Time, linear then quadratic interaction effects, and the constant. The *F*-contrast shown,

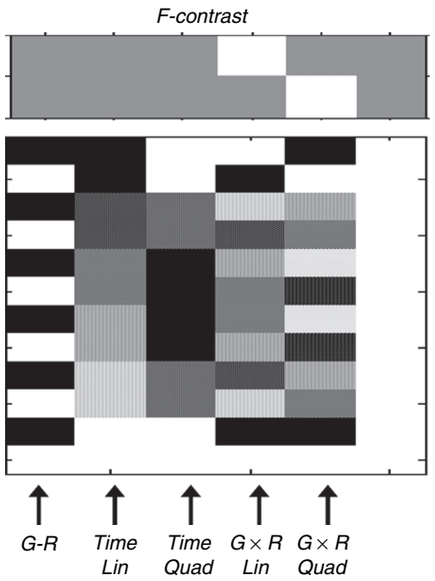

FIGURE 15.1 A single-subject design matrix and *F*-contrast showing non-linear (linear + quadratic) interactions in a 2 × 6 factorial design.

which picks out the fourth and fifth columns, would test for any type of time-by-condition interaction up to second order. Note that another common example of an interaction between a categorical factor and a parametric factor arises in psychophysiological interactions (PPIs; Chapter 38): in these cases, the psychological factor is often categorical (e.g. attended versus unattended) and the physiological factor is invariably parametric, since it reflects the continuous signal sampled by each scan from the source region of interest.

EVENT-RELATED fMRI, AND RANDOMIZED VERSUS BLOCKED DESIGNS

Event-related fMRI is simply the use of fMRI to detect responses to individual trials, in a manner analogous to the time-locked event-related potentials (ERPs) recorded with electroencephalography (EEG). The neuronal activity associated with each trial is normally (though not necessarily) modelled as a delta function – an 'event' – at the trial onset.

Historically, the advent of event-related methods (Dale and Buckner, 1997; Josephs *et al*., 1997; Zarahn *et al*., 1997), based on linear convolution models (see Chapter 14), offered several advantages. Foremost was the ability to intermix trials of different types (so-called 'randomized designs'), rather than blocking them in

the manner required for positron emission tomography (PET) and initially adopted for fMRI (so-called 'blocked designs'). The counterbalancing or randomizing of different trial-types, as is standard in behavioural or electrophysiological studies, ensures that the average response to a trial-type is not biased by a specific context or history of preceding trial-types. This is important because the blocking of trial-types might, for example, induce differences in the cognitive 'set' or strategies adopted by subjects. Johnson *et al*. (1997) for example, provided direct evidence that the presentation format – randomized or blocked – can affect the ERP associated with a trial-based memory effect.

Note that there are also disadvantages associated with randomized designs. Foremost, such designs are generally less efficient for detecting effects than are blocked designs (with short SOAs and reasonable block lengths; see below). In addition, some psychological manipulations, such as changes in selective attention or task, may be more appropriate when blocked.

Other advantages of event-related methods include:

1 the *post hoc* categorization of trial-types according to the subject's behaviour (e.g. Henson *et al*., 1999b), or *post hoc* parametric modulation of neuronal activity by reaction time (RT) for each trial
2 modelling events whose occurrence is beyond experimental control, such as those that can only be indicated by the subject (e.g. perceptual transitions in the face-evase illusion, Kleinschmidt *et al*., 1998)
3 the use of 'oddball' designs, in which the stimulus of interest is one that deviates from the prevailing context, and which therefore cannot be blocked (e.g. Strange *et al*., 2000).

Epochs versus events and state- versus item-effects

It is important to distinguish between the experimental design (randomized versus blocked) and the neuronal model (events versus epochs). For example, a blocked design can be modelled as a series of events. Indeed, modelling the BOLD response to each stimulus within a block may capture variability that is not captured by a simple 'epoch' (or boxcar) model, particularly for SOAs of more than a few seconds, which will lead to small fluctuations of the BOLD response around the mean 'block' response (Price *et al*., 1999; Mechelli *et al*., 2003a; see, e.g. Figure 15.2 bottom left).

In SPM, the choice of events versus epochs can also have important conceptual consequences. Consider, for example, an experiment with two blocks of words presented at different rates (once every 4 s versus once every 2 s). The data may be such that mean activity during the block of words presented at the fast rate may be greater, but not twice as great, as that for the slow rate. When modelling both conditions as epochs (upper panels of Figure 15.2), the parameter estimates for the two rates may be, for example, 3 and 5 respectively. If individual words were modelled as events, however (lower panels of Figure 15.2), the *relative* size of the parameter estimates could be reversed, e.g. 11 and 9 respectively. This is simply because the parameter estimates have different interpretations for the two types of model: in the epoch model, they reflect the response *per block*, whereas in the event model, they reflect the response *per word*. Since there are twice as many words in the fast- relative to slow-rate blocks, and yet the mean block activity is not double, the response per word must be less (i.e. a non-linear saturation as a function of word rate).

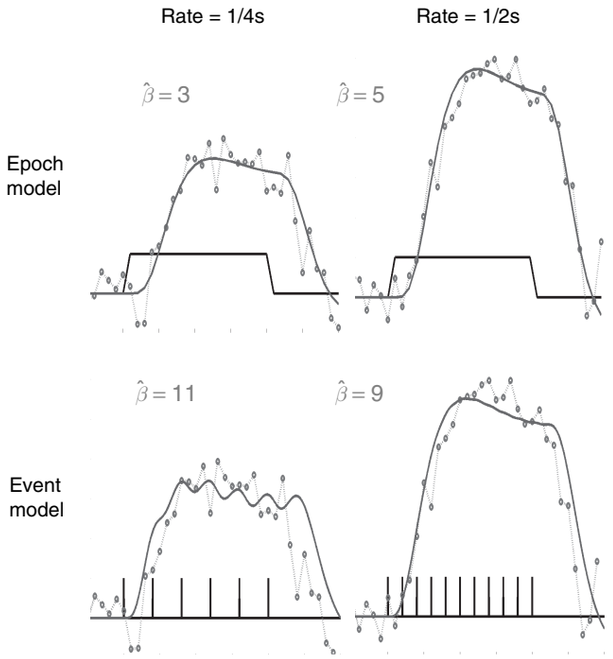

FIGURE 15.2 Effects of modelling the same data with events or epochs.

Another situation where this issue arises concerns trials of different duration. If all trials are of the same duration (and that duration is below ~2 s), then they can be modelled effectively as events because, after convolution with the haemodynamic response function (HRF), a difference in the duration of a trial causes a difference in the scaling of the predicted response, but has little effect on its shape (see Chapter 14). Since it is the scaling of the predicted response that is estimated in the GLM, changing the duration of all trials (from approx 0 to 2 s) simply changes the size of the resulting parameter estimates,

but has no effect on statistics.[1] For longer duration trials, the response begins to plateau, meaning that an 'epoch model' can be a better model. More important, however, is the case of trials that vary in duration from trial to trial within a condition, or across conditions. Whether these are better modelled as events, or as epochs of different durations (e.g. with duration equal to the RT for each trial), is debatable. For example, if the stimulus duration were constant and only RTs varied, then the activity in V1 may not be expected to vary with RT, so an event model might fit better (and in this case, the parameter estimate can be interpreted as the response *per trial*). For activity in premotor cortex on the other hand, greater activity might be expected for trials with longer RTs, so a 'varying-duration' epoch model might fit better (and in this case, the parameter estimate can be interpreted as the response *per unit time*). So the choice of model depends on the assumptions about the duration of neuronal activity in the particular region of interest. If this is unknown, trials whose durations vary over a few seconds (as with typical RTs) are probably best modelled with two regressors: one modelling events, and a second modelling a parametric modulation of the response, by the RT on each trial.

Finally, note that one can combine both events and epochs within the same model. A common example of this is when trying to distinguish between sustained ('state') effects and transient ('item') effects. Chawla *et al*. (1999), for example, investigated the interaction between selective attention (a state-effect) and transient stimulus changes (an item-effect) in such a 'mixed epoch-event' design. Subjects viewed a visual stimulus that occasionally changed in either colour or motion. In some blocks, they detected the colour changes, in other blocks they detected the motion changes. By varying the interval between changes within a block, Chawla *et al*. were able to reduce the correlation between the corresponding epoch- and event-related regressors (which increases the statistical efficiency to detect either effect alone; see below). Tests of the epoch-related effect showed that attending to a specific visual attribute (e.g. colour) increased the baseline activity in regions selective for that attribute (e.g. V4). Tests of the event-related effect showed that the impulse response to the same change in visual attribute was augmented when subjects were attending to it (Plate 18). These combined effects of selective attention – raising endogenous baseline activity and increasing the gain of the exogenous response – could not be distinguished in a blocked or fully randomized design.

Timing issues

There are two practical issues concerning the timing within randomized designs (which also apply to blocked designs, but to a lesser extent): the effective sampling rate of the BOLD response, and the different acquisition times for different slices within a scan (i.e. volume) when using echo-planar imaging (EPI).

It is possible to sample the impulse response at post-stimulus intervals, T_S, shorter than the inter-scan interval, T_R, by dephasing event onsets with respect to scan onsets (Josephs *et al*., 1997). This uncoupling can be effected by ensuring the SOA is not a simple multiple of the T_R, or by adding a random trial-by-trial delay in stimulus onsets relative to scan onsets (Figure 15.3). In both cases, responses at different peristimulus times (PST) are sampled over trials. The main difference between the two methods is simply whether the SOA is fixed or random, i.e. whether or not the stimulus onset is predictable. For example, an effective PST sampling of 0.5 Hz can be achieved with an SOA of 6 s and a T_R of 4 s; or by adding a delay of 0 or 2 s randomly to each trial (producing SOAs of 4–8 s, with a mean of 6 s). While effective sampling rates higher than the T_R do not necessarily improve response detection (since there is little power in the canonical response above 0.2 Hz), higher sampling rates are important for quantifying the response shape, such as its latency (Miezin *et al*., 2000; Henson and Rugg, 2001).

Dephasing event onsets with respect to scan onsets does not help the second practical issue concerning different slice acquisition times. This 'slice-timing' problem (Henson *et al*., 1999a) refers to the fact that, with a descending EPI sequence for example, the bottom slice is acquired T_R seconds later than the top slice. If a single basis function (such as a canonical HRF) were used to model the response, and onset times were specified relative to the start of each scan, the data in the bottom slice would be systematically delayed by T_R seconds relative to the model.[2] This would produce poor (and biased) parameter estimates for later slices, and mean

[1] This is despite the fact that the 'efficiency', as calculated by Eqn. **15.1**, increases with greater scaling of the regressors. This increase is correct, in the sense that a larger signal will be easier to detect in the presence of the same noise, but misleading in the sense that it is the size of the signal that we are estimating with our model (i.e. the data are unaffected by how we model the trials).

[2] One solution would be to allow different event onsets for different slices. However, slice-timing information is usually lost as soon as images are re-sliced relative to a different orientation (e.g. during spatial normalization).

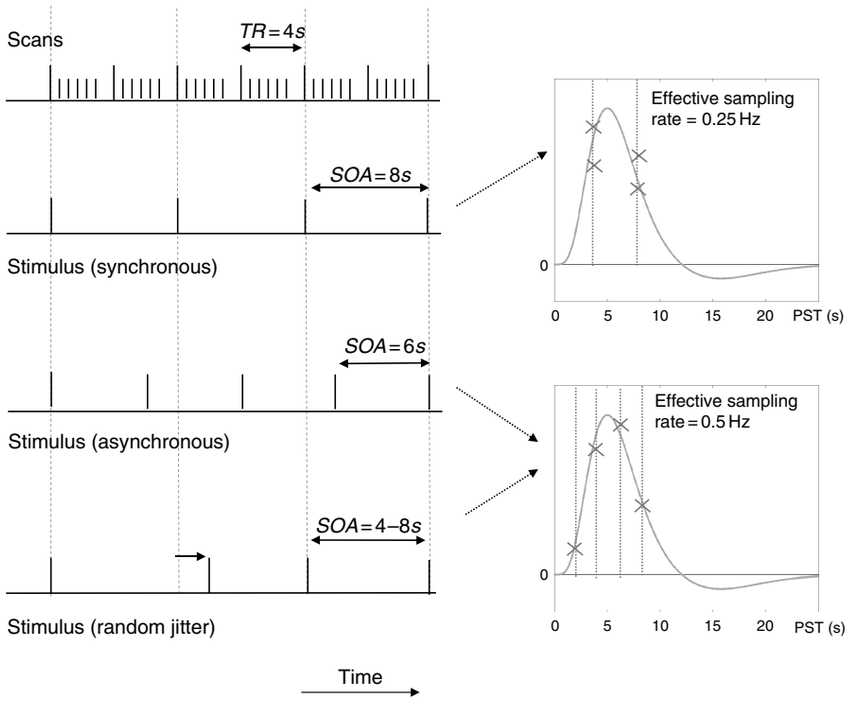

FIGURE 15.3 Effective sampling rate. Schematic (left) of event onsets relative to scan onsets (tall vertical lines represent first slice per scan; shorter lines represent subsequent slices) and resulting peristimulus sampling points (right).

that different sensitivities would apply to different slices (Figure 15.4(a)). There are two main solutions to this problem: to interpolate the data during pre-processing to make it seem as if the slices were acquired simultaneously; or use a temporal basis set that allows different response onset latencies.

Temporal interpolation of the data (using a full Fourier interpolation) is possible during pre-processing of images in SPM. One question that often arises is whether such temporal realignment should be performed before or after spatial realignment, given that movement often occurs. The answer depends on the order that slices are acquired within each scan. For sequential (contiguous) slice-acquisition, temporal interpolation is probably better performed after spatial realignment. This is because nearby voxels in space are sampled close in

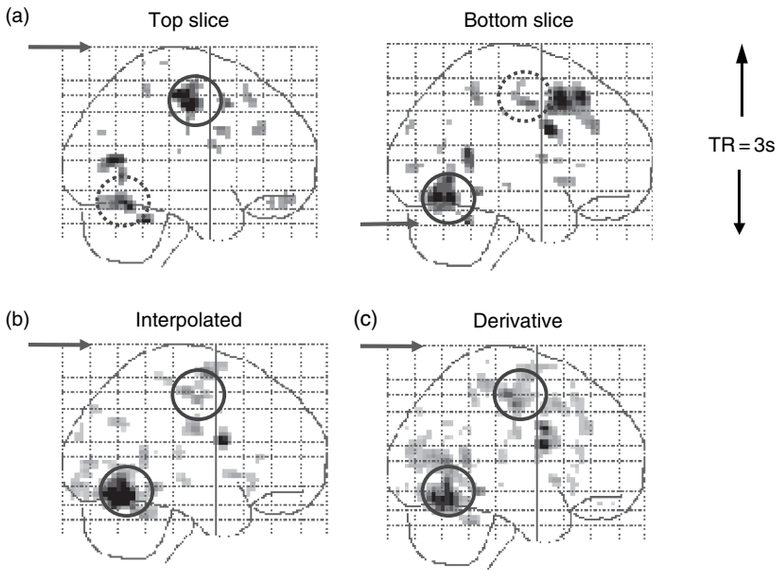

FIGURE 15.4 The slice-timing problem (from Henson *et al.*, 1999a) for a TR of 3 s. (a) SPM{*t*} for a [1] contrast on a canonical HRF synchronized with the top slice (left) or synchronized with the bottom slice (right). Note increased sensitivity to visual regions in latter case, but reduced sensitivity to motor regions. (b) SPM{*t*} when the model is synchronized with the top slice, but the data have been interpolated as if all slices were acquired at the time of the top slice. Note sensitivity recovered in both motor and visual regions. (c) SPM{*F*} for the canonical HRF and its temporal derivative. Note sensitivity again recovered in both motor and visual regions.

time. Therefore, the temporal error for a voxel whose signal comes from different acquisition slices, due to re-slicing after correction for movement across scans, will be small (given that movement is rarely more than a few 'slices-worth'). The alternative, of performing temporal realignment before spatial realignment could cause greater error, particularly for voxels close to boundaries with large signal differences (e.g. the edge of the cortex): in such cases, rapid movement may cause the same voxel to sample quite different signal intensities across successive scans. Such high-frequency changes are difficult to interpolate (temporally in this case). The order of preprocessing is not so clear for interleaved slice-acquisition schemes, in which adjacent slices can be sampled $\frac{1}{2}T_R$ seconds apart. In this case, and when there is no rapid movement, it may be better to perform temporal realignment before spatial realignment.

During slice-time correction, the data are interpolated by an amount proportional to their sampling time relative to a reference slice (whose data are unchanged). The event onsets can then be synchronized with the acquisition of that reference slice. In SPM, this is equivalent to maintaining event onsets relative to scan onsets, but setting the time-point T_0 in the simulated time-space of N time bins, from which the regressors are sampled (see Chapter 14), to $T_0 = round(nN/S)$ where the reference slice is the nth slice acquired of the S slices per scan. This can ameliorate the slice-timing problem, if one wishes to use a single assumed response form (e.g. canonical HRF, see Figure 15.4(b)). A problem with slice-timing correction is that the interpolation will alias frequencies above the Nyquist limit $1/(2T_R)$. Ironically, this means that the interpolation accuracy decreases as the slice-timing problem (i.e. T_R) increases. For short $T_R < 2$–3 s, the interpolation error is likely to be small. For longer T_R, the severity of the interpolation error may depend on whether appreciable signal power exists above the Nyquist limit (which is more likely for rapid, randomized event-related designs).

An alternative solution to the slice-timing problem is to include additional temporal basis functions (see Chapter 14) to accommodate the timing errors within the GLM. The Fourier basis set, for example, does not have a slice-timing problem (i.e. it is phase-invariant). For more constrained sets, the addition of the temporal derivative of the response functions may be sufficient (see Figure 15.4(c)). The parameter estimates for the derivatives will vary across slices, to capture shifts in the data relative to the model, while those for the response functions can remain constant (up to a first-order Taylor approximation, Chapter 14). The temporal derivative of the canonical HRF, for example, can accommodate slice-timing differences of approximately plus or minus a second, or a T_R up to 2 s (when the model is synchronized to the middle slice in time). A potential problem with this approach occurs when the true impulse responses are also shifted in time relative to the assumed response functions: the combined latency shift may exceed the range afforded by the temporal derivatives.

EFFICIENCY AND OPTIMIZATION OF fMRI DESIGNS

This section is concerned with optimizing experimental fMRI designs for a specific contrast of interest. The properties of the BOLD signal measured by fMRI – particularly the 'sluggish' nature of the impulse response and the presence of low-frequency noise – can make the design of efficient experiments difficult to intuit. This section therefore starts with some general advice, before explaining the reasons for this advice from the perspectives of:

1 signal-processing
2 statistical 'efficiency'
3 correlations among regressors.

General points

Scan for as long as possible

This advice is of course conditional on the subject being able to perform the task satisfactorily in a sustained fashion. Longer is better because the power of a statistical inference depends primarily on the degrees of freedom (*df*), and the *df* depend on the number of scans. One might therefore think that reducing the T_R (inter-scan interval) will also increase your power. This is true to a certain extent, though the 'effective' *df* depend on the temporal autocorrelation of the sampled data (i.e. 100 scans rarely means 100 independent observations; Chapter 14), so there is a limit to the power increase afforded by a shorter T_R.

If you are only interested in group results (e.g. extrapolating from a random sample of subjects to a population), then the statistical power normally depends more heavily on the number of subjects than the number of scans per subject (Friston *et al.*, 2002). In other words, you are likely to have more power with 100 scans on 20 subjects, than with 400 scans on 5 subjects, particularly given that inter-subject variability tends to exceed inter-scan variability. Having said this, there are practical issues, like the preparation time necessary to position the subject in the scanner, that mean that 100 scans on 20 subjects takes more time than 400 scans on 5 subjects. A common

strategy is therefore to run several experiments on each subject while they are in the scanner.

Keep the subject as busy as possible

This refers to the idea that 'dead-time' – time during which the subject is not engaged in the task of interest – should be minimized. Again, of course, there may be psychological limits to the subject's performance (e.g. they may need rests), but apart from this, factors such as the SOA should be kept as short as possible (even within blocks of trials). The only situation where you might want longer SOAs (or blocks of rest) is if you want to measure 'baseline'. From a cognitive perspective though, baseline is rarely meaningful, since it is rarely under strong experimental control (see below).

Only stop the scanner – i.e. break your experiment into sessions – if it is strictly necessary. Breaks in scanning disrupt the spin equilibrium (i.e. require extra dummy scans), reduce the efficiency of any temporal filtering (since the data no longer constitute a single time-series), and introduce other potential 'session' effects (McGonigle *et al.*, 2000).

Do not contrast trials that are remote in time

One problem with fMRI is that there is a lot of low-frequency noise. This has various causes, from aliased biorhythms to gradual changes in physical parameters (e.g. ambient temperature). Thus, any low-frequency 'signal' (induced by your experiment) may be difficult to distinguish from background noise. This is why SPM recommends a highpass filter (see Chapter 14). Since contrasts between trials that are far apart in time correspond to low-frequency effects, they may be filtered out.

In SPM, for example, a typical highpass cut-off is $1/128\,\text{s} \sim 0.01\,\text{Hz}$, based on the observation that the amplitude as a function of frequency, f, for a subject at rest has a '$1/f+$ white noise' form (Plate 19), in which amplitude reaches a plateau for frequencies above approximately 0.01 Hz (the inflection point of the '1/f' and 'white' noise components). When summing over frequencies (in a statistical analysis), the removal of frequencies below this cut-off will increase the signal-to-noise ratio (SNR), provided that most of the signal is above this frequency.

In the context of blocked designs, the implication is not to use blocks that are too long. For two alternating conditions, for example, block lengths of more than 50 s would cause the majority of signal (i.e. that at the fundamental frequency of the square-wave alternation) to be removed when using a highpass cut-off of 0.01 Hz. In fact, the optimal block length in an on-off design, regardless of any highpass filtering, is approximately 16 s (see below).

Randomize the order, or SOA, of trials close together in time

As will be explained below, in order to be sensitive to differences between trials close together in time (e.g. less than 20 s), one either uses a fixed SOA but varies the order of different trial-types (conditions), or constrains their order but varies the SOA. Thus, a design in which two trials alternate every 4 s is inefficient for detecting the difference between them. One could either randomize their order (keeping the SOA fixed at 4 s), or vary their SOA (keeping the alternating order).[3]

Signal-processing perspective

We begin by assuming that one has an event-related design, and the interest is in detecting the presence (i.e. measuring the amplitude) of a BOLD impulse response whose shape is well-characterized (i.e. a canonical HRF).[4] Given that we can treat fMRI scans as time-series, some intuition can be gained from adopting a signal-processing perspective, and by considering a number of simple examples.

To begin with, consider an event every 16 s. The result of convolving delta functions representing the events with the canonical HRF is shown in Figure 15.5(a) (see Chapter 14 for a discussion of linear convolution models). Maximizing the efficiency of a design is equivalent to maximizing the 'energy' of the predicted fMRI time-series, i.e. the sum of squared signal values at each scan (equal to the variance of the signal, after mean-correction). In other words, to be best able to detect the

[3] Note that, in this context, blocks can be viewed as runs of trials of the same type, and a blocked design corresponds to a varying-SOA design in which there is bimodal distribution of SOAs: a short SOA corresponding to the SOA within blocks, and a long SOA corresponding to the SOA between the last trial of one block and the first of the next.

[4] A distinction has been made between the ability to detect a response of known shape, 'detection efficiency' (as considered here), and the ability to estimate the shape of a response, 'estimation efficiency' [0] (Liu *et al.*, 2001;[**15.1] Birn *et al.*, 2002). This distinction actually reduces simply to the choice of temporal basis functions: The same efficiency equation (Eqn.**15.2** below) can be used to optimize either detection or estimation efficiency by using different response functions: e.g. either a canonical HRF or a FIR basis set respectively. A blocked design will optimize detection efficiency; whereas a randomized design with null events will optimize estimation efficiency (see Henson, 2004 for further details). Hagberg *et al.* (2001) considered a range of possible SOA distributions (bimodal in the case of blocked designs, exponential in the case of fully randomized designs) and showed that 'long-tail' distributions combine reasonable detection and estimation efficiency.

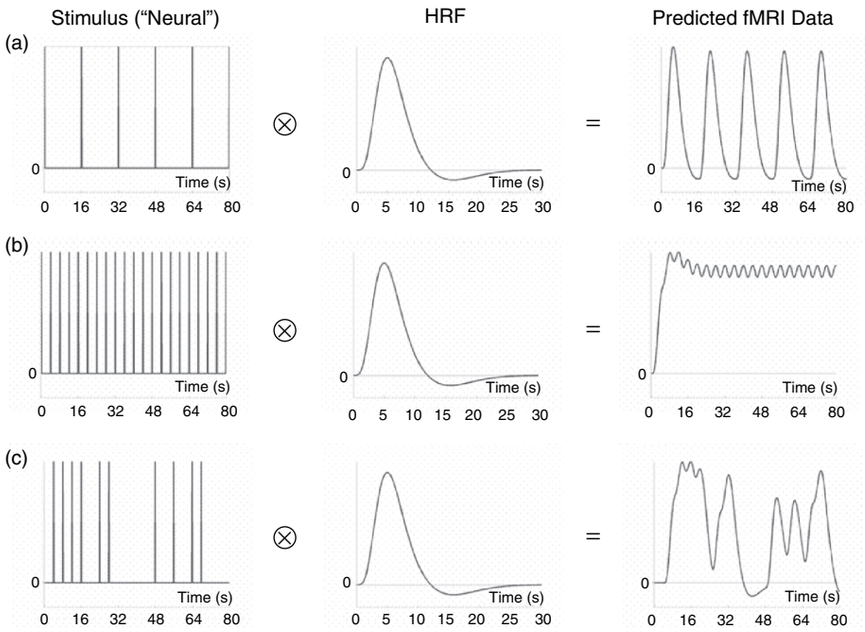

FIGURE 15.5 Effect of convolution by an assumed HRF on neuronal activity evoked by (a) events every 16 s, (b) events every 4 s and (c) events occurring with a 50 per cent probability every 4 s.

signal in the presence of background noise, we want to maximize the variability of that signal. A signal that varies little will be difficult to detect.

The above example (a fixed SOA of 16 s) is not particularly efficient, as we shall see later. What if we present the stimuli much faster, say every 4 s? The result is shown in Figure 15.5(b). Because the responses to successive events now overlap considerably, we see an initial build-up (transient) followed by small oscillations around a 'raised baseline'. Although the overall signal is high, its variance is low, and the majority of stimulus energy will be lost after highpass filtering (particularly after removal of the mean, i.e. lowest frequency). So this is an even less efficient design.

What if we vary the SOA randomly? Let's say we have a minimal SOA of 4 s, but only a 50 per cent probability of an event every 4 s. This is called a stochastic design (and one way to implement it is to intermix an equal number of 'null events' with 'true events'; see next section). This is shown in Figure 15.5(c). Though we only use half as many stimuli as in Figure 15.5(b), this is a more efficient design. This is because there is a much larger variability in the signal.

We could also vary the SOA in a more systematic fashion. We could have runs of events, followed by runs of no (null) events. This corresponds to a blocked design. For example, we could have blocks of 5 stimuli presented every 4 s, alternating with 20 s of rest, as shown in Figure 15.6(a). This is even more efficient than the previous stochastic design. To see why, we shall consider the Fourier transform of these time-series. First, however, note that, with short SOAs, the predicted fMRI time-series for a blocked design is similar to what would obtain if neuronal activity were sustained throughout the block (i.e. during the ISI [**15.2] as well) as in an epoch model (Figure 15.6(b)). Now, if we take the Fourier transform of each function in Figure 15.6(b), we can plot amplitude (magnitude) as a function of frequency (Figure 15.6(c). The amplitude spectrum of the square-wave stimulus function has a dominant frequency corresponding to its 'fundamental' frequency ($Fo = 1/(20\text{s} + 20\text{s}) = 0.025\,\text{Hz}$), plus a series of 'harmonics' ($3Fo$, $5Fo$, ... *etc*) of progressively decreasing amplitude. The fundamental frequency corresponds to the frequency of a sinusoidal that best matches the basic on-off alternation; the harmonics can be thought of as capturing the 'sharper' edges of the square-wave function relative to this fundamental sinusoid.

The reason for performing the Fourier transform is that it offers a slightly different perspective. Foremost, a convolution in time is equivalent to a multiplication in frequency space. In this way, we can regard the stimulus function as our original data and the HRF as a 'filter'. One can see immediately from the shape of the Fourier transform of the HRF that this filter will 'pass' low frequencies, but attenuate higher frequencies (this is why it is sometimes called a 'low-pass filter' or 'temporal smoothing kernel'). This property is why, for example, much high-frequency information was lost with the fixed SOA of 4 s in Figure 15.5(b). In the present example, the result of multiplying the amplitude spectrum of the stimulus function by that of the filter is that some of

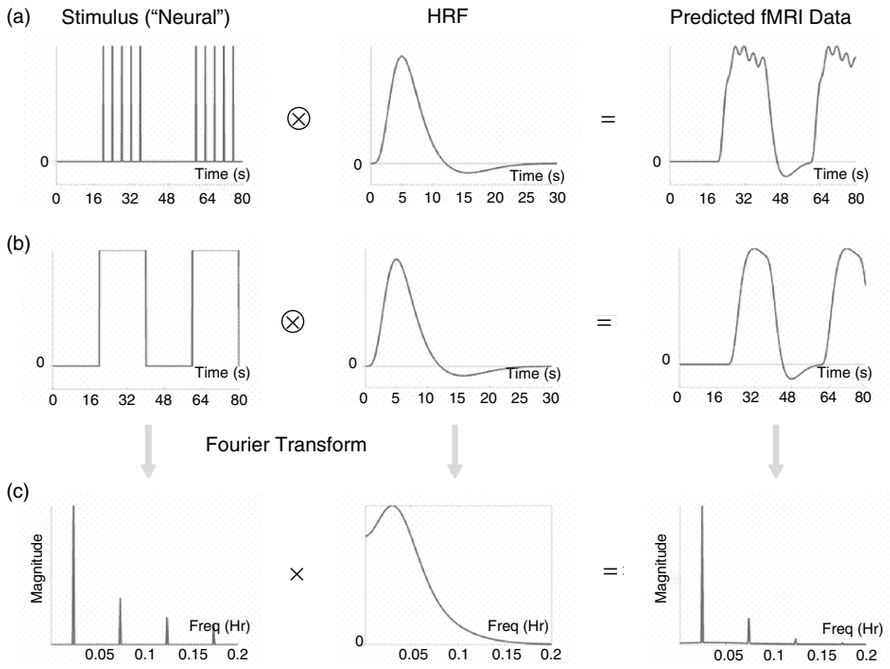

FIGURE 15.6 Effect of convolution by an assumed HRF on neuronal activity evoked by (a) blocks of events, (b) epochs of 20 s, and (c) the amplitude spectra after Fourier transform of (b).

the higher-frequency harmonics are attenuated, but the amplitude of the fundamental frequency is not. In other words, the majority of the signal is 'passed' by the HRF filter.

We are now in a position to answer the question: what is the most efficient design of all? Well, assuming we had a limited amount of total 'stimulus energy', the optimal design would be to modulate the neuronal activity in a sinusoidal fashion, with a frequency that matches the peak of the amplitude spectrum of the HRF filter. With the canonical HRF used here, this would be ~0.03 Hz (1/30 s). The sinusoidal modulation places all the stimulus energy at this single frequency, shown by the single line in frequency space in Figure 15.7.

We can now also turn to the question of highpass filtering. Because the filtering is commutative, we can apply the highpass filter to the low-pass filter inherent in the HRF to create a single band-pass filter (or 'effective HRF', Josephs and Henson, 1999). This is shown in Figure 15.8, in which the highpass filter reflects the 'chunk' of low frequencies that has been removed from the HRF filter (highpass cut-off here = 1/120 s ~0.008 Hz). The consequence of highpass filtering is shown for long blocks of 80 s (20 trials every 4 s). Because the fundamental

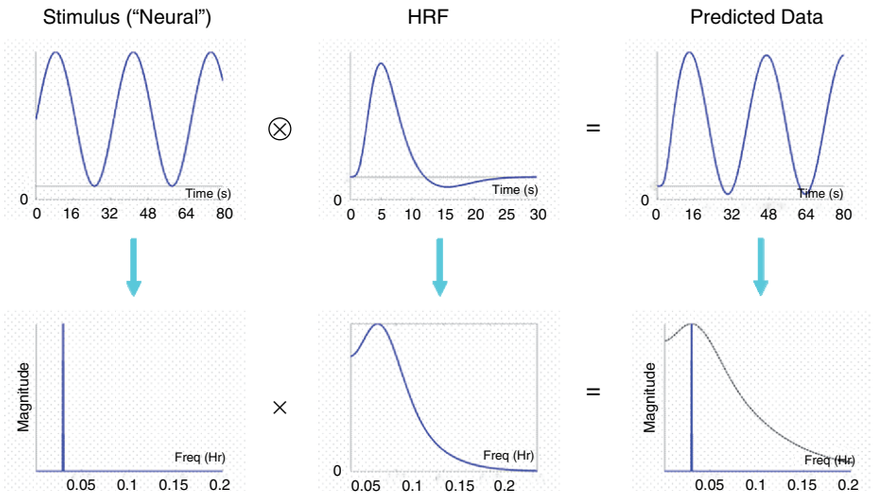

FIGURE 15.7 Effect of convolution by an assumed HRF on sinusoidal neuronal activity of 0.03 Hz.

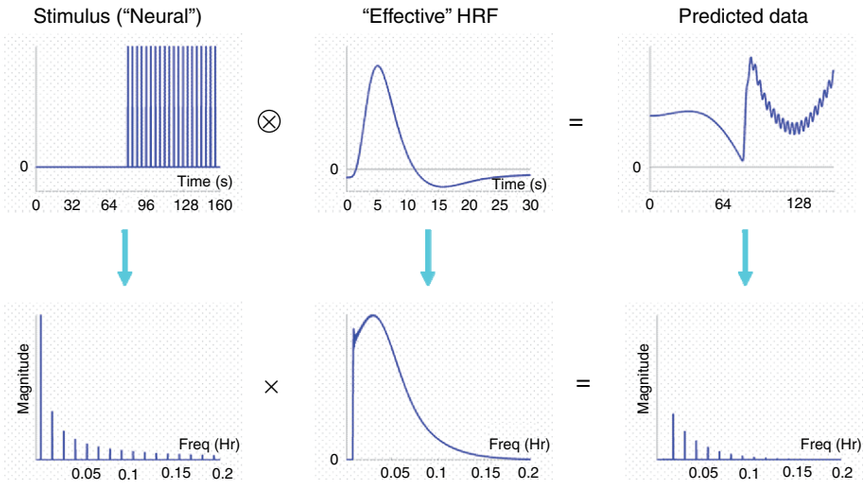

FIGURE 15.8 Effect of convolution by an 'effective' HRF (i.e. including a high-pass filter) on 80-s blocks of trials.

frequency in this design (1/160 s ~0.006 Hz) is lower than the highpass cut-off, a large proportion of signal energy is lost (reflected by the rather strange shape of the predicted fMRI time-series, in which the lowest frequency has been removed). This is therefore not an efficient design (with this specific highpass cut-off). This illustrates the general point that blocked designs are only efficient when the block length is not too long: approx 15 s-on, 15 s-off is optimal (see Figure 15.7). Block durations of up to 50 s-on, 50 s-off are also fine (given that the HRF filter does not attenuate low frequencies much), but block durations much longer than this (or contrasts between two of many different types of 50 s-blocks) may be in danger of being swamped by low-frequency noise.

Finally, we can return to consider what happens in a stochastic design like that in Figure 15.5(c). The effect of the randomized SOA is to 'spread' the signal energy across a range of frequencies, as shown in Figure 15.9. Some of the high- and low-frequency components are lost to the effective HRF filter, but much is passed, making it a reasonably efficient design.

Statistical perspective

From the statistical perspective, the aim is to minimize the standard error of a *t*-contrast, $c^T\hat{\beta}$ (i.e. the denominator of a *t*-statistic, Chapter 8). Given the specified contrast of interest, c, and parameter estimates, $\hat{\beta}$, the variance of $c^T\hat{\beta}$ is given by (Friston *et al.*, 2000):

$$\mathrm{var}(c^T\hat{\beta}) = \sigma^2 c^T (SX)^+ SVS^T (SX)^{+T} c \qquad \textbf{15.2}$$

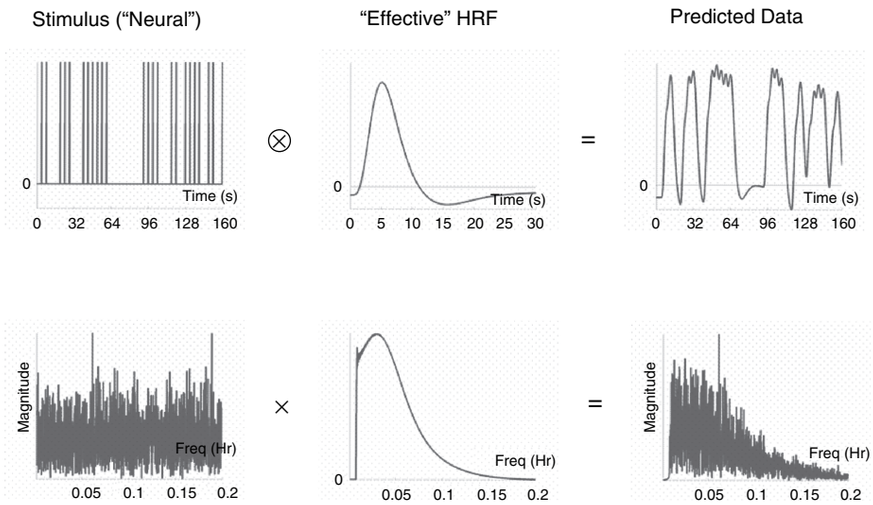

FIGURE 15.9 Effect of convolution by an 'effective' HRF on randomized SOA events (minimum = 4 s).

where S is a filter matrix incorporating the highpass filter and any temporal smoothing, and V is the noise autocorrelation matrix. We want to *minimize* this variance with respect to the design matrix, X. If we assume that the filter matrix S is specified appropriately to 'whiten' the residuals, such that $SVS^T = I$ (i.e. when $S = K^{-1}$, where $KK^T = V$; Chapter 14), and we incorporate S into X, then this is equivalent to *maximizing* the efficiency, ζ:

$$\xi(\sigma^2, c, X) = (\sigma^2 c^T (X^T X)^{-1} c)^{-1} \quad \textbf{15.3}$$

For a given contrast, c, this equation can be split into the 'noise variance', σ^2, and the 'design variance', $(X^T X)^{-1}$ (Mechelli *et al*., 2003b). If one assumes that the noise variance is independent of the specific design used (which may not be the case, Mechelli *et al*., 2003b; see later), then the efficiency of a contrast for a given design is proportional to:

$$\xi(c, X) = (c^T (X^T X)^{-1} c)^{-1} \quad \textbf{15.4}$$

(For F-contrasts, where c is a matrix, the *trace* operator can be used to reduce efficiency to a single number; Dale, 1999). Note that $\zeta(c, X)$ has no units; it is a relative measure. It depends on the scaling of the design matrix and the scaling of the contrasts. Thus, all we can really say is that one design is more efficient than another (for a given contrast). In what follows, we use Eqn.**15.4** to evaluate the efficacy of different sorts of design and look at how designs can be characterized probabilistically.

Stochastic designs

For a single event-type, the space of possible experimental designs can be captured by two parameters: the minimal SOA (Δt) and the probability, p_t, of an event occurring at every Δt (Friston *et al*., 1999). In 'deterministic' designs, $p_t = 1$ or $p_t = 0$, giving a series of events with fixed SOA, as in Figure 15.5(a). In 'stochastic' designs $0 \leq p_t \leq 1$, producing a range of SOAs (as in Figure 15.5(c)). For 'stationary' stochastic designs, p_t is constant, giving an exponential distribution of SOAs; for 'dynamic' stochastic designs, p_t changes with time. The extreme case of a dynamic stochastic design is one in which the temporal modulation of p_t conforms to a square-wave, corresponding to a blocked design. Notice that the quantities p_t and Δt parameterize a space of design matrices probabilistically. In other words, they specify the probability $p(X|p_t, \Delta t)$ of getting any particular design matrix. This allows one to compute the expected design efficiency for any class that can be parameterized in this way:

$$\langle \xi(c, p_t, \Delta t) \rangle = \int p(X|p_t, \Delta t) \xi(c, X) dX \quad \textbf{15.5}$$

This expected design efficiency can be evaluated numerically by generating large number of design matrices (using p_t and Δt) and taking the average efficiency according to Eqn. **15.4**. Alternatively, one can compute the expected efficiency analytically as described in Friston *et al*. (1999). This allows one to explore different sorts of designs by treating the design matrix itself as a random variable. For stochastic designs, efficiency is generally maximal when the Δt is minimal and the (mean) $p_t = 1/(L+1)$, where L is the number of trial types (see Friston *et al*., 1999).

Figure 15.10 shows the expected efficiency for detecting canonical responses to a single event-type versus baseline, i.e. $L = 1$ and $c = 1$, for a range of possible designs. The deterministic design with $\Delta t = 8$ s (top row) is least efficient, whereas the dynamic stochastic design with a square-wave modulation with $\Delta t = 1$ s is the most efficient (corresponding, in this case, to a 32 s on-off blocked design). Intermediate between these extremes are the dynamic stochastic designs that use a sinusoidal modulation of p_t. In other words, these designs produce clumps of events close together in time, interspersed with periods in which events are rarer. Though such designs are less efficient than the blocked (square-wave) design, they are more efficient than the stationary stochastic design with $\Delta t = 1$ s (second row of Figure 15.10), and assuming that subjects are less likely to notice the 'clumping' of events (relative to a fully blocked design), may offer a good compromise between efficiency and subjective unpredictability.

Transition probabilities

The space of possible designs can also be characterized by Δt and a 'transition matrix' (Josephs and Henson, 1999). This is a generalization of the above formulation that introduces conditional dependencies over time. For $L > 1$ different event-types, a $L^m \times L$ transition matrix captures the probability of an event being of each type, given the history of the last m event-types. A fully randomized design with two event-types (A and B) has a simple first-order transition matrix in which each probability is a half. The efficiencies of two contrasts, the main effect of A and B (versus baseline), $c = [1\ 1]^T$, and the differential effect, $c = [1\ -1]^T$, are shown as a function of Δt in Plate 20(a). The optimal SOA for the main effect under these conditions is approximately 20 s. The efficiency of the main effect decreases for shorter SOAs, whereas the efficiency of the differential effect increases. Clearly, the efficiency for the differential contrast cannot increase indefinitely as the SOA decreases; at some point, the BOLD response must saturate (see below). Nonetheless, this graph clearly

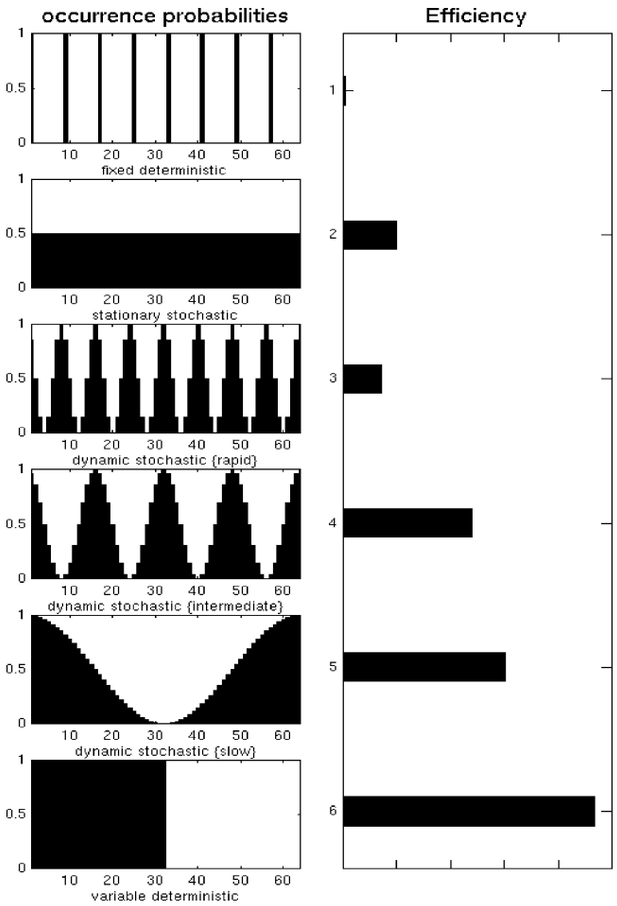

FIGURE 15.10 Efficiency for a single event-type (from Friston *et al*., 1999). (a) Probability of event each SOA (left column) and expected design efficiency (right column, increasing left-to-right) for a deterministic design with $\Delta t = 8$ s (1st row), a stationary stochastic (randomized) design with $p_t = 0.5$ (2nd row) and dynamic stochastic designs with modulations of p_t by different sinusoidal frequencies (3rd to 5th rows) and in a blocked manner every 32 s (6th row).

demonstrates how the optimal SOA depends on the specific contrast of interest.[5]

Various experimental constraints on multiple event-type designs can also be considered. In some situations, the order of event-types might be fixed, and the design question relates to the optimal SOA. For a design in which A and B must alternate (e.g. where A and B are transitions between two perceptual states), the optimal SOA for a differential effect is 10 s (Plate 20(b), i.e. half of that for the main effect). In other situations, experimental constraints may limit the SOA, to at least 10 s say, and the design question relates to the optimal stimulus ordering. An alternating design is more efficient than a randomized design for such intermediate SOAs. However, an alternating design may not be advisable for psychological reasons (subjects' behaviour might be influenced by the predictable pattern). In such cases, a permuted design (in which each of trial-types is presented successively in a randomly-permuted order) may be a more suitable choice (see Plate 20(b)).

A further design concept concerns 'null events'. These are not real events, in that they do not differ from the baseline and hence are not detectable by subjects (so are not generally modelled in the design matrix). They were introduced by Dale and Buckner (1997) as 'fixation trials', to allow 'selective averaging' (see Chapter 14). In fact, they are simply a convenient means of creating a stochastic design by shuffling a certain proportion of null events among the events of interest (producing an exponential distribution of SOAs). From the perspective of multiple event-type designs, the reason for null events is to buy efficiency for both the main effect and differential effect at short SOAs (at a slight cost to the efficiency for the differential effect; see Plate 20(c)).

The efficiencies shown in Plate 20 are unlikely to map simply (e.g. linearly) onto the size of the *t*-statistic. Nonetheless, if the noise variance, in Eqn. **15.3**, is independent of experimental design, the relationship should at least be monotonic. Mechelli *et al*. (2003b) showed that the noise variance can vary significantly between a blocked and a randomized design (both modelled with events). This suggests that the stimulus ordering did affect (un-modelled) psychological or physiological effects in this dataset, contributing to the residual error. When the data were highpass filtered however, the noise variance no longer differed significantly between the two designs. In this case, the statistical results were in agreement with the relative efficiencies predicted from the estimation variances.

Efficiency in terms of correlations

Another way of thinking about efficiency is in terms of the correlation between (contrasts of) regressors within the design matrix. In Eqn. **15.3** the term X^TX is called the information matrix and reflects the orthogonality of the design matrix. High covariance between the columns of the design matrix introduces redundancy. This can increase the covariance of the parameter estimates $(X^TX)^{-1}$ and lead to low efficiency (depending on the particular contrast).

Consider the earlier example of two event-types, A and B, randomly intermixed, with a short SOA. If we plot

[5] The main effect, which does not distinguish A and B, is of course equivalent to a deterministic design, while the differential effect is equivalent to a stochastic design (from the perspective of any one event-type).

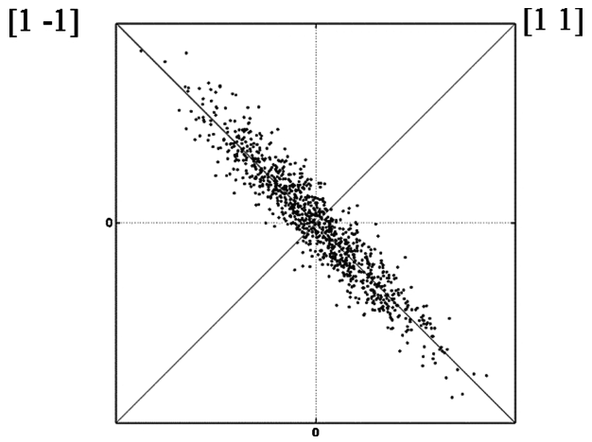

FIGURE 15.11 Scatter plot for two mean-corrected regressors (one point per scan) corresponding to two event-types randomly intermixed with a short SOA.

the resulting two regressors (after convolution with an HRF) against each other, we would end up with a scatter plot something like that in Figure 15.11, where each point reflects one scan. The high negative correlation between the regressors is because whenever there is high signal for A, there tends to be low signal for B, and vice versa. If we consider the projection of this distribution onto the $x = -y$ direction (corresponding to a [1 −1] contrast), it will have a large dispersion, i.e. high experimental variance, which means the difference between A and B will be detected efficiently in this design. However, if we project the distribution onto the $x = y$ direction (corresponding to a [1 1] contrast), it will have little spread, i.e. low experimental variance, which means that we will not detect the common effect of A and B versus baseline, efficiently. This demonstrates the markedly different efficiency for these two contrasts at short SOAs that was shown in Plate 20(a).

Projection onto the x or y axes (i.e. [1 0] or [0 1] contrasts) will have less spread than if the two regressors were orthogonal and formed a spherical cloud of points. This shows how correlations can reduce efficiency and makes an important general point about correlations. High correlation between two regressors means that the parameter estimate for each one will be estimated inefficiently, i.e. the parameter estimate itself will have high variance. In other words, if we estimated each parameter many times we would get wildly different results. In the extreme case, that the regressors are perfectly correlated, the parameters would be inestimable (i.e. they would have infinite variance). Nonetheless, we could still estimate efficiently the difference between them. Thus, high correlations within the orthogonality matrix shown by SPM should not be a cause of concern for some contrasts: what is really relevant is the correlation between the *contrasts* of interest (i.e. linear combinations of columns of the design matrix) relative to the rest of the design matrix (i.e., null space of the contrast).

In short-SOA, randomized designs with no null events, for example, we might detect brain regions showing a reliable difference between event-types, yet when we plot the event-related response, we might find they are all 'activations' versus baseline, all 'deactivations' versus baseline or some activations and some deactivations. However, these impressions are more apparent than real (and should not really be shown). If we tested the reliability of these activations or deactivations, they are unlikely to be significant. This is because we cannot estimate the baseline reliably in such designs. This is why, for such designs, it does not make sense to plot error bars showing the variability of each condition alone: one should plot error bars pertaining to the variability of the difference (i.e. that of the contrast actually tested).

Orthogonalizing

Another common misapprehension is that one can overcome the problem of correlated regressors by 'orthogonalizing' one regressor with respect to the other. This rarely solves the problem. The parameter estimates always pertain to the orthogonal part of each regressor (this is an automatic property of fitting within the GLM). Thus, neither the parameter estimate for the orthogonalized regressor, nor its variance, will change. The parameter estimate for the other regressor will change. However, this parameter estimate now reflects the assumption that the common variance is uniquely attributed to this regressor. We must have an *a priori* reason for assuming this (i.e. without such prior knowledge, there is no way to determine which of the two correlated regressors caused the common effect). In the absence of such knowledge, there is no reason to orthogonalize.

The conception of efficiency in terms of correlations can help with the design of experiments where there is necessarily some degree of correlation among regressors. Two main experimental situations where this arises are:

1 when trials consist of two events, one of which must follow the other
2 blocks of events in which one wishes to distinguish 'item-' from 'state-' effects (see above).

A common example of the first type of experiment are 'working memory' designs, in which a trial consists of a stimulus, a short retention interval, and then a response. We shall ignore the retention interval and concentrate on how one can separate effects of the stimulus from those of the response. With short SOAs between each event-type (e.g. 4 s), the regressors for the stimulus and response will be negatively correlated, as shown in Figure 15.12(a). Two possible solutions to this problem are shown in

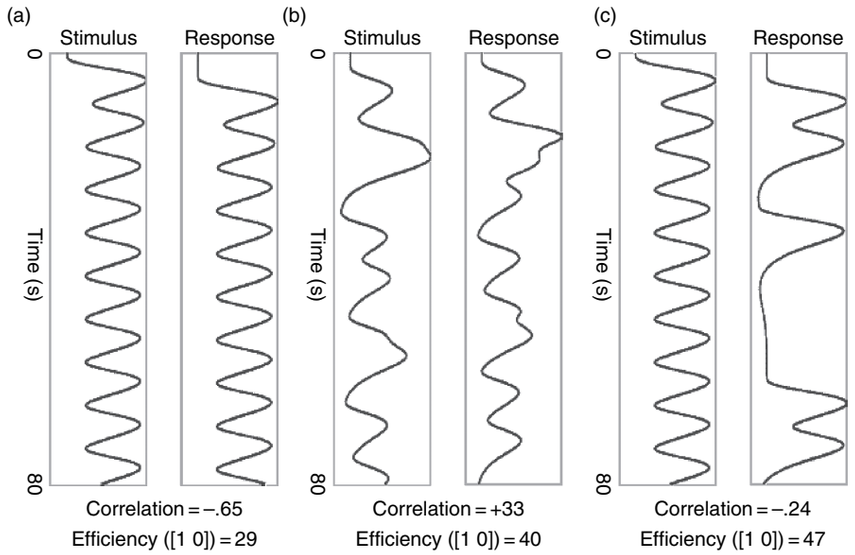

FIGURE 15.12 Regressors for 'working memory' trials presented every 8 s, consisting of (a) stimulus followed after 4 s by a response, (b) stimulus-response intervals varied from 0 to 8 s, and (c) responses following stimuli by 4 s, but only on 50 per cent of trials.

Figure 15.12(b) and 15.12(c). The first is to vary the time between successive stimuli and responses (assuming this is possible and that this variation is large; e.g. 1–8 s). The second is to keep the stimulus-response interval fixed at 4 s, but only cue a response on a random half of trials. The effect of both is to reduce the correlation between the regressors, which increases the efficiency separate brain activity related to stimuli from that related to responses.

The second type of experiment tries to distinguish transient responses (item-effects) from sustained responses (state-effects). Such separation of transient and sustained effects requires modelling blocks of trials in terms of both individual events within blocks and sustained epochs throughout the blocks. An example with a fixed SOA of 4 s between events is shown in Figure 15.13(a). Here, the correlation between the event and epoch regressors is naturally high, and the efficiency for detecting either effect alone is low. Using the same total number of events per block, but with a pseudo-randomized design in which the events are randomly spread over the block with a minimal SOA of 2 s (Figure 15.13(b)), the correlation is reduced and efficiency increased. (Note that one perverse consequence of having to introduce some long SOAs between events within blocks in such 'mixed designs' is that subjects may be less able to maintain a specific cognitive 'state'.)

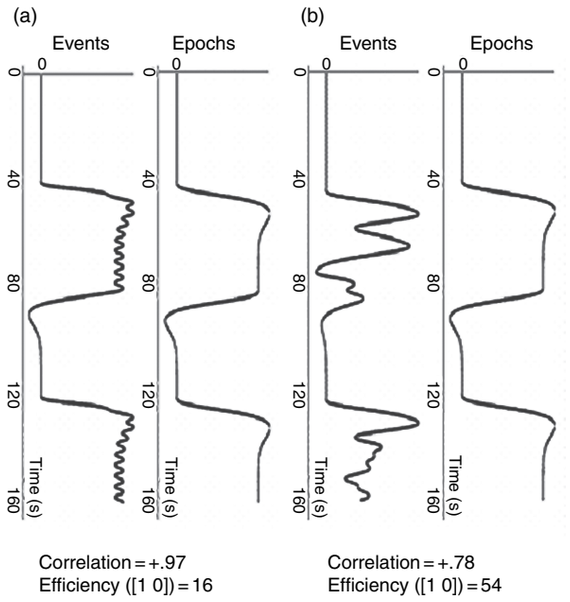

FIGURE 15.13 Regressors for 'mixed designs' that attempt to separate transient (item) from sustained (state) effects. (a) 10 events per block presented every SOA of 4 s, (b) 10 events per block distributed randomly over 2-s SOAs.

Effect of non-linearities on efficiency

The above efficiency arguments have assumed linearity, i.e. that the responses to successive trials summate linearly, no matter how close together in time they occur. In reality, we know there is a 'saturation', or under-additivity, even for SOAs of about 10 s (see Chapters 14 and 27). This means that the efficiency for stochastic designs does not increase indefinitely as

the SOA decreases (e.g. for the differential effect in Plate 20(a)). By estimating non-linearity with a Volterra expansion (Chapter 14), Friston *et al*. (1998) calculated the impact of such non-linearity on evoked responses. The result is shown in the insert in Plate 20(a). The dotted line shows the average response to a train of stimuli under linear assumptions; the solid line shows the effects of saturation (using a second-order Volterra kernel). While the solid line is below the dotted line for all SOAs (below 10 s), the divergence is small until SOAs of 1–2 s. Indeed, the prediction of this calculation is that the optimal SOA can be as low as 1 s, i.e. the advantage of short SOAs can outweigh the saturation of responses until surprisingly short SOAs (though it should be noted that this prediction is based on a specific dataset, and may not generalize). Indeed, differential responses between randomized event-types have been detected with SOAs as short as 0.5 s (Burock *et al*., 1998).

In this chapter, we have looked at how to detect evoked fMRI responses efficiently. Before turning to models of evoked responses in EEG in the next chapter, we will conclude with some common questions that exercise people designing fMRI studies

COMMON QUESTIONS

What is the minimum number of events I need?

Unfortunately, there is no answer to this, other than 'the more, the better'. The statistical power depends on the effect size and variability, and this is normally unknown. Heuristics like 'you cannot do an event-related fMRI analysis with less than N events' are fairly meaningless, unless one has a specific effect size in mind (which is likely to be a function of the brain region, the scanner strength, the sequence type, etc.). Note it is possible that fewer trials are required (for a given power) than for an equivalent contrast of behavioural data (e.g. if the noise level in, say, RTs exceeds that in a specific cortical region contributing to those RTs). Furthermore, it is not even the number of events *per se* that is relevant, it is also the SOA and event-ordering (see next question).

Do shorter SOAs mean more power simply because there are more trials?

It is not simply the number of trials: the temporal deployment of those trials is vital (as explained above). Thus 400 stimuli every 3 s is *less* efficient than 40 stimuli every 30 s for detecting a single event-related response (since a fixed SOA of 3 s produces little experimental variability after convolution by the HRF). Two hundred stimuli occurring with a 50 per cent probably every 3 s (i.e. pseudo-randomly mixed with 200 null events) is much more efficient than either.

What is the maximum number of conditions I can have?

A common interpretation of the rule – do not compare trials that are too far apart in time – is not to design experiments with too many experimental conditions. More conditions necessarily mean that replications of a particular condition will be further apart in time. However, the critical factor is not the number of conditions *per se*, but the specific contrasts performed over those conditions. For pair-wise comparisons of only two of, say, eight blocked conditions the above caveat would apply: if there were equal numbers of blocks of each condition, blocks longer than 12.5 s (100 s/8) are likely to entail a substantial loss of signal when using a high-pass cut-off of 0.01 Hz. However, this caveat would not apply if the contrasts of interest included (i.e. 'spanned') all eight conditions. This would be the case if the experimenter were only interested in the two main effects and the interaction within a 2×4 factorial design (i.e. contrasts like $[1\ 1\ 1\ 1\ -1\ -1\ -1\ -1]$). If you must compare or plot only a subset of many such blocked conditions, you should consider presenting those blocks in a fixed order, rather than random or counterbalanced order, which will minimize the time between replications of each condition, i.e. maximize the frequency of the contrast.

Should I use null events?

Null events are simply a convenient means of achieving a stochastic distribution of SOAs, in order to allow estimation of the response versus inter-stimulus baseline, by randomly intermixing them with the events of interest. However, the 'baseline' may not always be meaningful. It may be well defined for V1, in terms of visual flashes versus a dark background. It becomes less well defined for 'higher' regions associated with cognition because it is unclear what these regions are 'doing' during the inter-stimulus interval. The experimenter normally has little control over this. Moreover, the baseline does not control for the fact that the events of interest are impulsive (rapid changes), whereas the baseline is sustained (and may entail adaptation or reduced attention). For this reason, it is often better to forget about baseline and add an extra low-level control event instead.

Another problem with null events is that, if they are too rare (e.g. less than approximately 33 per cent), they actually become 'true' events in the sense that subjects may be expecting an event at the next SOA and so be surprised when it does not occur (the so-called 'missing stimulus' effect that is well-known in event-related potential (ERP) research). One solution is to replace randomly intermixed null events with periods of baseline between runs of events (i.e. 'block' the baseline periods). This will increase the efficiency for detecting the common effect versus baseline, at a slight cost in efficiency for detecting differences between the randomized event-types within each block. Yet another problem is that the unpredictability of the occurrence of true events (caused by the randomly intermixed null events) can cause delayed or even missed processing of the events of interest, if subjects cannot prepare for them.

In summary, null events are probably only worthwhile if:

1 you think the mean activity during the constant inter-stimulus interval is meaningful to contrast against
2 you do not mind null events being reasonably frequent (to avoid 'missing stimulus' effects)
3 you do not mind the stimulus occurrence being unpredictable (as far as the subject is concerned).

Having said this, some form of baseline can often serve as a useful 'safety net' (e.g. if you fail to detect differences between two visual event-types of interest, you can at least examine V1 responses and check that you are seeing a basic evoked response to both event-types – if not, you can question the quality of your data or accuracy of your model). Moreover, you may need randomly to inter-mix null events if you want to estimate more precisely the shape of the BOLD impulse response (see footnote 4). It is often the case that people include a low-level baseline or null event to use as reference for a localizing contrast on tests for differences among true events. In other words, the contrast testing for all events versus baseline can serve as a useful constraint on the search volume for interesting comparisons among events.

Should I generate multiple random designs and choose the most efficient?

This is certainly possible, though be wary that such designs are likely to converge on designs with some structure (e.g. blocked designs, given that they tend to be optimal, as explained above). This may be problematic if such structure affects subjects' behaviour (particularly if they notice the structure). Note, however, that there are software tools available that optimize designs at the same time as allowing users to specify a certain level of counterbalancing (to avoid fully blocked designs, e.g. Wager and Nichols, 2003).

REFERENCES

Birn RM, Cox RW, Bandettini PA (2002) Detection versus estimation in event-related fMRI: choosing the optimal stimulus timing. *NeuroImage* **15**: 252–64

Burock MA, Buckner RL, Woldorff MG *et al*. (1998) Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *NeuroReport* **9**: 3735–39

Chawla D, Rees G, Friston KJ (1999) The physiological basis of attentional modulation in extrastriate visual areas. *Nat Neurosci* **2**: 671–76

Dale AM (1999) Optimal experimental design for event-related fMRI. *Hum Brain Mapp* **8**: 109–14

Dale A, Buckner R (1997) Selective averaging of rapidly presented individual trials using fMRI. *Hum Brain Mapp* **5**: 329–40

Friston KJ, Price CJ, Fletcher P *et al*. (1996) The trouble with cognitive subtraction. *NeuroImage***4**: 97–104

Friston KJ, Josephs O, Rees G *et al*. (1998) Non-linear event-related responses in fMRI. *Mag Res Med* **39**: 41–52

Friston KJ, Zarahn E, Josephs O *et al*. (1999) Stochastic designs in event-related fMRI. *NeuroImage* **10**: 607–19

Friston KJ, Josephs O, Zarahn E *et al*. (2000) To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. *NeuroImage* **12**: 196–208

Friston KJ, Glaser DE, Henson RNA *et al*. (2002) Classical and Bayesian inference in neuroimaging: applications. *NeuroImage* **16**: 484–512

Friston KJ, Penny WD, Glaser DE (2005) Conjunction revisited. *NeuroImage* [15.3]

Hagberg GE, Zito G, Patria F *et al*. (2001) Improved detection of event-related functional MRI signals using probability functions. *NeuroImage* **14**: 193–205

Henson RNA (2004) Analysis of fMRI time-series: linear time-invariant models, event-related fMRI and optimal experimental design. In *Human Brain Function*, 2nd edn, Frackowiak RS, Friston KJ, Frith CD *et al*. (eds). Elsevier, London, pp 793–822

Henson RNA (2005) What can functional neuroimaging tell the experimental psychologist? *Quart J Exp Psych A* **58**: 193–234

Henson RNA, Büchel C, Josephs O *et al*. (1999a) The slice-timing problem in event-related fMRI. *NeuroImage* **9**: 125

Henson RNA, Rugg MD, Shallice T *et al*. (1999b) Recollection and familiarity in recognition memory: an event-related fMRI study. *J Neurosci* **19**: 3962–72

Henson RNA, Rugg MD (2001) Effects of stimulus repetition on latency of the BOLD impulse response. *NeuroImage* **13**: 683

Johnson MK, Nolde SF, Mather M *et al*. (1997) Test format can affect the similarity of brain activity associated with true and false recognition memory. *Psychol Sci* **8**: 250–57

Josephs O, Turner R, Friston KJ (1997) Event-related fMRI. *Hum Brain Mapp* **5**: 243–48

Josephs O, Henson RNA (1999) Event-related fMRI: modelling, inference and optimisation. *Phil Trans Roy Soc Lond* **354**: 1215–28

Kleinschmidt A, Büchel C, Zeki S *et al*. (1998) Human brain activity during spontaneously reversing perception of ambiguous figures. *Proc R Soc Lond B Biol Sci* **265**: 2427–33

McGonigle DJ, Howseman AM, Athwal BS *et al*. (2000) Variability in fMRI: an examination of intersession differences. *NeuroImage* **11**: 708–34

Mechelli A, Henson RNA, Price CJ *et al*. (2003a) Comparing event-related and epoch analysis in blocked design fMRI. *NeuroImage* **18**: 806–10

Mechelli A, Price CJ, Henson RNA *et al*. (2003b) The effect of high-pass filtering on the efficiency of response estimation: a comparison between blocked and randomised designs. *NeuroImage* **18**: 798–805

Miezin FM, Maccotta L, Ollinger JM *et al*. (2000) Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *NeuroImage* **11**: 735–59

Nichols TE, Brett M, Andersson J *et al*. (2005) Valid conjunction inference with the minimum statistic. *NeuroImage* **25**: 653–60

Price CJ, Friston KJ (1997) Cognitive conjunction: a new approach to brain activation experiments. *NeuroImage* **5**: 261–70

Price CJ, Veltman DJ, Ashburner J *et al*. (1999) The critical relationship between the timing of stimulus presentation and data acquisition in blocked designs with fMRI. *NeuroImage* **10**: 36–44

Sternberg S (1969) The discovery of processing stages: extensions of Donders method. *Acta Psychol* **30**: 276–315

Strange BA, Henson RN, Friston KJ *et al*. (2000) Brain mechanisms for detecting perceptual, semantic, and emotional deviance. *NeuroImage* **12**: 425–33

Wager TD, Nichols TE (2003) Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *NeuroImage* **18**: 293–309

Zarahn E, Aguirre G, D'Esposito M (1997) A trial-based experimental design for fMRI. *NeuroImage* **6**: 122–38