Detection Power, Estimation Efficiency, and Predictability in Event-Related fMRI

Thomas T. Liu,* Lawrence R. Frank,*,† Eric C. Wong,*,‡ and Richard B. Buxton*

*Department of Radiology and ‡Department of Psychiatry, University of California, San Diego, La Jolla, California 92037; and
†Veterans Administration San Diego Healthcare System, La Jolla, California 92037

Experimental designs for event-related functional magnetic resonance imaging can be characterized by both their detection power, a measure of the ability to detect an activation, and their estimation efficiency, a measure of the ability to estimate the shape of the hemodynamic response. Randomized designs offer maximum estimation efficiency but poor detection power, while block designs offer good detection power at the cost of minimum estimation efficiency. Periodic single-trial designs are poor by both criteria. We present here a theoretical model of the relation between estimation efficiency and detection power and show that the observed trade-off between efficiency and power is fundamental. Using the model, we explore the properties of semirandom designs that offer intermediate trade-offs between efficiency and power. These designs can simultaneously achieve the estimation efficiency of randomized designs and the detection power of block designs at the cost of increasing the length of an experiment by less than a factor of 2. Experimental designs can also be characterized by their predictability, a measure of the ability to circumvent confounds such as habituation and anticipation. We examine the relation between detection power, estimation efficiency, and predictability and show that small increases in predictability can offer significant gains in detection power with only a minor decrease in estimation efficiency. © 2001 Academic Press

INTRODUCTION

Event-related experimental designs for functional magnetic resonance imaging (fMRI) have become increasingly popular because of their flexibility and their potential for avoiding some of the problems, such as habituation and anticipation, of more traditional block designs (Buckner *et al.,* 1996, 1998; Dale and Buckner, 1997; Josephs *et al.,* 1997; Zarahn *et al.,* 1997; Burock *et al.,* 1998; Friston *et al.,* 1998a, 1999; Rosen *et al.,* 1998; Dale, 1999; Josephs and Henson, 1999). In the evaluation of the sensitivity of experimental designs, it

is useful to distinguish between the ability of a design to detect an activation, referred to as detection power, and the ability of a design to characterize the shape of the hemodynamic response, referred to as estimation efficiency (Buxton *et al.,* 2000). Stimulus patterns in which the interstimulus intervals are properly randomized from trial to trial achieve optimal estimation efficiency (Dale, 1999) but relatively low detection power. Block designs, in which individual trials are tightly clustered into "on" periods of activation alternated with "off" control periods, obtain high detection power but very poor estimation efficiency. Dynamic stochastic designs have been proposed as a compromise between random and block designs (Friston *et al.,* 1999). These designs regain some of the detection power of block designs, while retaining some of the ability of random designs to reduce preparatory or anticipatory confounds.

In this paper we present a theoretical model that describes the relation between estimation efficiency and detection power. With this model we are able to show that the trade-off between estimation efficiency and detection power, as exemplified by the difference between block designs and random designs, is in fact fundamental. That is, any design that achieves maximum detection power must necessarily have minimum estimation efficiency, and any design that achieves maximum estimation efficiency cannot attain the maximum detection power.

We also examine an additional factor that is often implicit in the decision to adopt random designs. This is the perceived randomness of a design. Regardless of considerations of estimation efficiency, randomness can be critical for minimizing confounds that arise when the subject in an experiment can too easily predict the stimulus pattern. For example, studies of recognition using familiar stimuli and novel stimuli are hampered if all of the familiar stimuli are presented together. We introduce predictability as a metric for the perceived randomness of a design and explore the relation between detection power, estimation efficiency, and predictability.

The structure of this paper is as follows. After a brief review of the general linear model in the context of fMRI experiments, we present definitions for estimation efficiency and detection power and derive theoretical bounds for both quantities. We then describe a simple model that relates estimation efficiency and detection power and explore how the model can be used to understand the performance of existing experimental designs and also to generate new types of designs. We next provide a definition for predictability and describe methods for measuring it. Simulation results are used to support the theoretical results and to clarify the trade-offs between detection power, estimation efficiency, and predictability.

THEORY

General Linear Model

The general linear model provides a flexible framework for analyzing fMRI signals (Friston *et al.,* 1995b; Dale, 1999). In matrix notation, we write the model as

$$\mathbf{y} = \mathbf{Xh} + \mathbf{Sb} + \mathbf{n}, \tag{1}$$

where $\mathbf{y}$ is a $N \times 1$ vector that represents the observed fMRI time series, $\mathbf{X}$ is a $N \times k$ design matrix, $\mathbf{h}$ is a $k \times 1$ parameter vector, $\mathbf{S}$ is a $N \times l$ matrix consisting of nuisance model functions, $\mathbf{b}$ is a $l \times 1$ vector of nuisance parameters, and $\mathbf{n}$ is a $N \times 1$ vector that represents additive Gaussian noise. We assume that the covariance of the noise vector $\mathbf{n}$ is given by $\mathbf{C_n} = \sigma^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix and σ^2 is an unknown variance term that needs to be estimated from the data.

In this paper, we focus on the case in which the columns of the design matrix $\mathbf{X}$ are shifted versions of a binary stimulus pattern consisting of 1's and 0's and the parameter vector $\mathbf{h}$ represents the hemodynamic response (HDR) that we wish to estimate. In other words, $\mathbf{Xh}$ is the matrix notation for the discrete convolution of a stimulus pattern with the hemodynamic response. For example, in the case in which the stimulus pattern is [1 0 1 1 0 0] and there are three parameters in the HDR, we have

$$\mathbf{y} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} + \mathbf{Sb} + \mathbf{n}.$$

In the following sections, we characterize the estimation efficiency and detection power obtained with different binary stimulus patterns. When there are Q event types and HDRs of interest, the design matrix

may be written as $\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2 \cdots \mathbf{X}_Q]$ and the parameter vector as $\mathbf{h} = [\mathbf{h}_1^T \ \mathbf{h}_2^T \cdots \mathbf{h}_Q^T]^T$, where each matrix $\mathbf{X}_i$ consists of shifted binary stimulus patterns for the ith event type and $\mathbf{h}_i$ is the vector for the corresponding HDR (Dale, 1999). In general, stimulus patterns need not be binary. The use of graded stimuli has proven to be useful in characterizing the response of various neural systems (Boynton *et al.,* 1996). For an event-related design a graded pattern might have the form [1 0 2.5 3.0 0 0]. The optimal design of graded stimulus patterns can be addressed within the theoretical framework presented here, but is beyond the scope of this paper.

The term $\mathbf{Sb}$ in the linear model represents nuisance effects that are of no interest, e.g., a constant term, linear trends, or low-frequency drifts. The columns of $\mathbf{S}$ are typically chosen to be low-frequency sine and cosine functions (Friston *et al.,* 1995a) or low-order Legendre polynomials. For most fMRI experiments, $\mathbf{S}$ should at the very least contain a constant term and a linear trend term, e.g., the zeroth- and first-order Legendre polynomials. Following Scharf and Friedlander (1994), we refer to the subspaces spanned by the columns of $\mathbf{X}$ and $\mathbf{S}$ as the *signal* subspace $\langle \mathbf{X} \rangle$ and the *interference* subspace $\langle \mathbf{S} \rangle$, respectively. These subspaces lie within the N-dimensional space spanned by the data. We require $\langle \mathbf{X} \rangle$ and $\langle \mathbf{S} \rangle$ to be linearly independent subspaces, so that no column in $\mathbf{X}$ can be expressed as a linear combination of the columns of $\mathbf{S}$ and vice versa. However, we do not require $\langle \mathbf{X} \rangle$ and $\langle \mathbf{S} \rangle$ to be orthogonal subspaces (i.e., there is no requirement that $\mathbf{S}^T \mathbf{X} = 0$), since this is too severe of a restriction. For example, most block designs are not orthogonal to linear trends. Finally, the space spanned by both $\mathbf{X}$ and $\mathbf{S}$ is denoted as $\langle \mathbf{XS} \rangle$.

Estimation Efficiency

A useful geometric approach to the problem of estimation in the presence of subspace interference has been described in Behrens and Scharf (1994) and serves as the basis of our analysis. The maximum likelihood estimate of $\mathbf{h}$ is written as

$$\hat{\mathbf{h}} = (\mathbf{X}^T \mathbf{P}_\mathbf{S}^\perp \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}_\mathbf{S}^\perp \mathbf{y}, \tag{2}$$

where $\mathbf{P}_\mathbf{S}^\perp = \mathbf{I} - \mathbf{S}(\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T$ is a projection matrix that removes the part of a vector that lies in the interference subspace $\langle \mathbf{S} \rangle$. In other words, $\mathbf{P}_\mathbf{S}^\perp$ removes nuisance effects such as linear trends. The estimate of the signal is $\mathbf{X}\hat{\mathbf{h}}$, which is the oblique projection $\mathbf{E_x y}$ of the data onto the signal subspace $\langle \mathbf{X} \rangle$, where $\mathbf{E_X} = \mathbf{X}(\mathbf{X}^T \mathbf{P}_\mathbf{S}^\perp \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}_\mathbf{S}^\perp$. A geometric picture of the oblique projection is shown in Fig. 1. It is important to note that, in general, the oblique projection is not the same as the projection of the data with interference terms
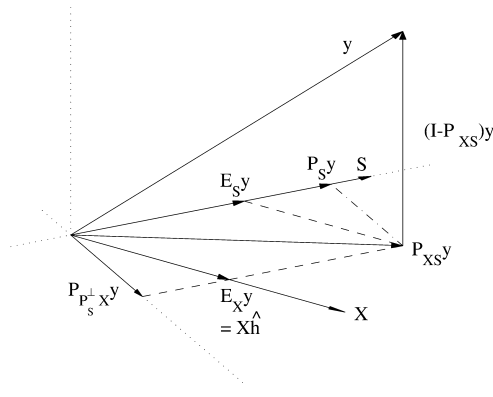
FIG. 1. Geometric picture of estimation and detection (adapted, by permission of the publisher, from Scharf and Friedlander, 1994; © 1994 IEEE). The data vector $\mathbf{y}$ is decomposed into a component, $\mathbf{P_{XS}y}$, that lies in the combined signal and interference subspace $\langle \mathbf{XS} \rangle$ and an orthogonal component $(\mathbf{I} - \mathbf{P_{XS}})\mathbf{y}$. The oblique projections of $\mathbf{y}$ onto the signal and interference subspaces are $\mathbf{E_X y}$ and $\mathbf{E_S y}$, respectively. The parameter estimate $\hat{\mathbf{h}}$ is the value of the parameter vector for which $\mathbf{X}\hat{\mathbf{h}}$ is equal to the oblique projection $\mathbf{E_X y}$. $\mathbf{P_{P_S^\perp X} y}$ is the projection of the data onto the part of $\mathbf{X}$ that is orthogonal to $\mathbf{S}$ and is equal to $\mathbf{P_{XS}y} - \mathbf{P_S y}$, where $\mathbf{P_S y}$ is the projection of the data onto $\mathbf{S}$. The F statistic is proportional to the ratio of the squared lengths of $\mathbf{P_{P_S^\perp X} y}$ and $(\mathbf{I} - \mathbf{P_{XS}})\mathbf{y}$. Note that while the estimation of the hemodynamic response does not require orthogonality of $\mathbf{S}$ and $\mathbf{X}$, the statistical significance, as gauged by the F statistic, is degraded when $\mathbf{S}$ and $\mathbf{X}$ are not orthogonal.

removed ($\mathbf{P_S^\perp y}$) onto the signal subspace $\langle \mathbf{X} \rangle$. That is, $\mathbf{X}(\mathbf{X}^T\mathbf{P_S^\perp X})^{-1}\mathbf{X}^T\mathbf{P_S^\perp y}$ does not equal $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{P_S^\perp y}$, unless $\langle \mathbf{X} \rangle$ and $\langle \mathbf{S} \rangle$ are orthogonal subspaces.

Equation (2) can be rewritten in the form

$$\hat{\mathbf{h}} = (\mathbf{X}_\perp^T\mathbf{X}_\perp)^{-1}\mathbf{X}_\perp^T\mathbf{y}, \quad (3)$$

where $\mathbf{X}_\perp = \mathbf{P_S^\perp X}$ is simply the design matrix with nuisance effects removed from each column. The covariance of the estimate is $\mathbf{C}_{\hat{\mathbf{h}}} = \sigma^2(\mathbf{X}_\perp^T\mathbf{X}_\perp)^{-1}$, and the sum of the variances of the components of $\hat{\mathbf{h}}$ is $\sigma^2\mathbf{trace}[(\mathbf{X}_\perp^T\mathbf{X}_\perp)^{-1}]$. The efficiency of the estimate can be defined as the inverse of the sum of the variances,

$$\xi = \frac{1}{\sigma^2\mathbf{trace}[(\mathbf{X}_\perp^T\mathbf{X}_\perp)^{-1}]}. \quad (4)$$

Experimental designs that maximize the estimation efficiency are referred to as *A-optimal* designs (Seber, 1977). The definition of estimation efficiency stated in Eq. (4) was introduced into the fMRI literature by Dale (1999) and serves as the starting point for our analysis.

Orthogonal Designs Maximize Estimation Efficiency

Estimation efficiency is maximized when $\mathbf{trace}[(\mathbf{X}_\perp^T\mathbf{X}_\perp)^{-1}]$ is minimized. It can be shown that this occurs when the columns of $\mathbf{X}_\perp$ are mutually

orthogonal (Seber, 1977). When there is only one event type, each column of $\mathbf{X}_\perp$ is obtained by first applying an appropriate shift to the binary stimulus pattern and then removing nuisance effects. The trace expression is therefore minimized with binary stimulus patterns, which, after detrending, are orthogonal to shifted versions of themselves.

In principle, orthogonality can be achieved by stimulus patterns that are realizations of a Bernoulli random process, which is the formal description of the random coin toss experiment. To generate a candidate stimulus pattern, we repeatedly flip a coin that has a probability P of landing "heads" and $1 - P$ of landing "tails," assigning a 1 to the stimulus pattern when we obtain heads and a 0 otherwise. The outcome of each toss is independent of the outcome of the previous toss. The binary stimulus pattern that we generate has two important properties. First, after removal of the mean value of the pattern (i.e., a constant nuisance term), the pattern is *on average* orthogonal to all possible shifts of itself. That is, the *expected value* of the inner product of the sequence with any shifted version is zero. Second, the pattern after removal of the mean is *on average* orthogonal to all other nuisance terms. This means that, aside from a constant nuisance term, the pattern is *on average* unaffected by the process of removing nuisance terms. As a result of these two properties, the design matrix $\mathbf{X}$ with columns that are shifted versions of a Bernoulli-type stimulus pattern results in a matrix $\mathbf{X}_\perp$ with columns that are *on average* orthogonal.

Bounds on Estimation Efficiency

Designs based on Bernoulli-type stimulus patterns are optimal in a statistical sense only, meaning that while *on average* they are optimal, some patterns may be suboptimal. A standard procedure is to generate a large number of random patterns and select the one with the best performance (Dale, 1999; Friston *et al.,* 1999). A theoretical upper bound on performance is useful in judging how good the "best" random pattern is.

To derive a bound on estimation efficiency, we first note that $\mathbf{trace}[(\mathbf{X}_\perp^T\mathbf{X}_\perp)^{-1}] = \sum_{i=1}^k 1/\lambda_i$, where λ_i is the ith eigenvalue of $\mathbf{X}_\perp^T\mathbf{X}_\perp$ (Seber, 1977). With any fixed value for the sum of the eigenvalues, the term $\sum_{i=1}^k 1/\lambda_i$ is minimized when all of the eigenvalues are equal. Since the sum of the eigenvalues is equal to $M = \mathbf{trace}[\mathbf{X}_\perp^T\mathbf{X}_\perp]$, we may write $\lambda_i = M/k$, which yields $\sum_{i=1}^k 1/\lambda_i = k^2/M$. If we assume that there are m 1's out of N total time points in the stimulus pattern and the constant term has been removed, then the energy of any one column of $\mathbf{X}_\perp$ is at most

$(1 - m/N)m$, where we define the energy of a vector as its magnitude squared. This leads directly to

$$M \leq (1 - m/N)mk. \qquad (5)$$

Placing the above results into Eq. (4), we obtain the bound

$$\xi \leq \frac{(1 - m/N)m}{k}, \qquad (6)$$

where we have assumed unit variance for the noise. The bound stated in Eq. (5) does not take into account the fact that for a random sequence with m 1's out of N total time points, the energy of shifted columns will decrease as more 1's are shifted out of the sequence. This effect slightly reduces the trace term M. An approximate bound on M that takes this effect into account is given in the Appendix and is used when comparing theoretical results to simulations.

The bound stated in Eq. (6) is maximized for the choice $m = N/2$, i.e., the number of 1's in the stimulus pattern is equal to half the number of total time points. This is consistent with the previously reported finding that, for the case of one event type, estimation efficiency is maximized when the probability of obtaining a 1 in the stimulus pattern is 0.5 (Friston *et al.,* 1999).

We should emphasize that the bound stated in Eq. (6) is specific to the case in which there is one event type. A full treatment of estimation efficiency for experiments with multiple event types is beyond the scope of this paper, but it is worth mentioning a few salient points. We assume that the stimulus patterns are mutually exclusive, meaning that, at each time point, at most one event type may have a 1 in its stimulus pattern. In addition, we assume that the probability P of obtaining a 1 is the same for all event types. With these assumptions and making use of the formalism described in Friston *et al.* (1999) for calculation of the expected value of $\mathbf{X}_\perp^T \mathbf{X}_\perp$, it can be shown that the maximum efficiency is in fact not obtained when the columns of $\mathbf{X}_\perp$ are orthogonal. Instead, the maximum efficiency is obtained for a probability of occurrence that achieves an optimal balance between two competing goals: (1) maximizing the energy in each of the columns of $\mathbf{X}_\perp$ and (2) reducing the correlation between columns. For two event types, this occurs for a probability $P = 1 - \sqrt{2}/2 = 0.29$, or equivalently, $m/N = 0.29$. An additional consideration that arises for multiple event types is the estimation efficiency for differences between event types. In order to equalize the efficiencies for both the individual event types and the differences, the optimal probability is $P = 1/(Q + 1)$, where Q is the number of event types (Burock *et al.,* 1998; Friston *et al.,* 1999).

Detection

The detection problem is formally stated as a choice between two hypotheses:

H_0, $\mathbf{y} = \mathbf{Sb} + \mathbf{n}$

(null hypothesis, no signal present), and

H_1, $\mathbf{y} = \mathbf{Xh} + \mathbf{Sb} + \mathbf{n}$

(signal present).

To decide between the two hypotheses, we compute an F statistic of the form

$$F = \frac{N - k - l}{k} \frac{\mathbf{y}^T \mathbf{P}_{\mathbf{P}_{\bar{\mathbf{S}}}^\perp \mathbf{x}} \mathbf{y}}{\mathbf{y}^T (\mathbf{I} - \mathbf{P_{XS}}) \mathbf{y}}, \qquad (7a)$$

where $\mathbf{P_{XS}}$ is the projection onto the subspace $\langle \mathbf{XS} \rangle$ and $\mathbf{P}_{\mathbf{P}_{\bar{\mathbf{S}}}^\perp \mathbf{x}} = \mathbf{P}_{\bar{\mathbf{S}}}^\perp \mathbf{X} (\mathbf{X}^T \mathbf{P}_{\bar{\mathbf{S}}}^\perp \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}_{\bar{\mathbf{S}}}^\perp$ is the projection onto the part of the signal subspace $\langle \mathbf{X} \rangle$ that is orthogonal to the interference subspace $\langle \mathbf{S} \rangle$ (Scharf and Friedlander, 1994). The F statistic is the ratio between an estimate $\mathbf{y}^T \mathbf{P}_{\mathbf{P}_{\bar{\mathbf{S}}}^\perp \mathbf{x}} \mathbf{y}/k$ of the average energy that lies in the part of the signal subspace $\langle \mathbf{X} \rangle$ that is orthogonal to $\langle \mathbf{S} \rangle$ and an estimate $\mathbf{y}^T (\mathbf{I} - \mathbf{P_{XS}}) \mathbf{y}/(N - k - l)$ of the noise variance σ^2 derived from the energy in the data space that is not accounted for by energy in the combined signal and interference subspace $\langle \mathbf{XS} \rangle$. Figure 1 provides a geometric interpretation of the quantities in Eq. (7a). As originally introduced into the fMRI literature by Friston *et al.* (1995b), the F statistic may also be written using the extra sum of squares principle (Draper and Smith, 1981) as

$$F = \frac{N - k - l}{k} \frac{\mathbf{y}^T (\mathbf{P_{XS}} - \mathbf{P_S}) \mathbf{y}}{\mathbf{y}^T (\mathbf{I} - \mathbf{P_{XS}}) \mathbf{y}}. \qquad (7b)$$

Equations (7a) and (7b) are equivalent, since $\mathbf{P}_{\mathbf{P}_{\bar{\mathbf{S}}}^\perp \mathbf{x}} = \mathbf{P_{XS}} - \mathbf{P_S}$ as can be verified upon inspection of Fig. 1.

When the null hypothesis H_0 is true, F follows a central F distribution with k and $N - k - l$ degrees of freedom. When hypothesis H_1 is true, F follows a noncentral F distribution with k and $N - k - l$ degrees of freedom and noncentrality parameter (Scharf and Friedlander, 1994),

$$\eta = \frac{\mathbf{h}^T \mathbf{X}^T \mathbf{P}_{\bar{\mathbf{S}}}^\perp \mathbf{X} \mathbf{h}}{\sigma^2}. \qquad (8)$$

The noncentrality parameter has the form of a signal-to-noise ratio in which the numerator is the expected energy of the signal after interference terms have been removed and the denominator is the expected noise variance.

To use the F statistic, we compare it to a threshold value β. If $F > \beta$, we choose hypothesis H_1 and declare that a signal is present; otherwise we choose the null hypothesis H_0. In most fMRI applications, the threshold β is chosen to achieve a desired probability of false alarm, i.e., the probability that we choose H_1 when H_0 is true. This probability can be computed from the central F distribution. Once the dimensions of $\mathbf{X}$ and $\mathbf{S}$ are known, the probability of false alarm is independent of $\mathbf{X}$ since the shape of the central distribution depends only on the dimensions k and $N - k - l$. As a result, all binary stimulus patterns of the same length yield the same probability of false alarm under the null hypothesis H_0, i.e., no activation. In practice, the dimension l of the interference subspace $\mathbf{S}$ is not known, although for most fMRI experiments l is typically between 1 and 5. Ignorance of l does not, however, alter the fact that only the dimension of $\mathbf{X}$, as opposed to its specific form, affects the probability of false alarm.

The probability of detection refers to the probability that we choose H_1 when H_1 is true and is also referred to as the *power* of a detector. For a given threshold value β, the detection power using the F statistic increases with the noncentrality parameter η. From Eq. (8), we can see that the noncentrality parameter depends directly on the design matrix $\mathbf{X}$. Once we have chosen β to achieve a desired probability of false alarm, we should select a design matrix that maximizes η. The noncentrality parameter is analogous to the estimated measurable power as defined by Josephs and Henson (1999).

In the degenerate case in which there is only one unknown parameter ($k = 1$), the F statistic is simply the square of the t statistic (Scharf and Friedlander, 1994). This typically corresponds to the situation in which we assume a known shape for the hemodynamic response function and are trying to estimate the amplitude of the activation. The detection power still depends on the noncentrality parameter as defined in Eq. (8), where $\mathbf{h}$ is the assumed known shape. To be explicit, if we rewrite the linear model as $\mathbf{y} = \mu\mathbf{z} + \mathbf{Sb} + \mathbf{n}$, where $\mathbf{z} = \mathbf{Xh}$ is the stimulus pattern convolved with the known shape (normalized to have unit amplitude) and μ is the unknown amplitude of the response, then the noncentrality parameter is $\eta = \mu^2\mathbf{z}^T\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{z}/\sigma^2 = \mu^2\mathbf{h}^T\mathbf{X}^T\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{Xh}/\sigma^2$.

Bounds on Detection Power

It is convenient to rewrite the noncentrality parameter as

$$\eta = \frac{\mathbf{h}^T\mathbf{X}_{\perp}^T\mathbf{X}_{\perp}\mathbf{h}}{\sigma^2}, \tag{9}$$

where $\mathbf{X}_{\perp}$ was defined previously as the design matrix $\mathbf{X}$ with nuisance effects removed from its columns. In determining the dependence of η on $\mathbf{X}_{\perp}$, we can ignore σ^2, which is just a normalizing factor over which we have no control. Furthermore, we normalize η by the energy $\mathbf{h}^T\mathbf{h}$ of the parameter vector $\mathbf{h}$ to obtain the Rayleigh quotient (Strang, 1980),

$$R = \frac{\mathbf{h}^T\mathbf{X}_{\perp}^T\mathbf{X}_{\perp}\mathbf{h}}{\mathbf{h}^T\mathbf{h}}. \tag{10}$$

The Rayleigh quotient can be interpreted as the noncentrality parameter obtained when the energy of the parameter vector and the variance of the noise are both equal to unity. It serves as a useful measure of the detection power of a given design.

The maximum of the Rayleigh quotient is equal to the maximum eigenvalue λ_1 of $\mathbf{X}_{\perp}^T\mathbf{X}_{\perp}$ and is attained when $\mathbf{h}$ is parallel to the eigenvector $\mathbf{v}_1$ associated with λ_1 (Strang, 1980). The maximum eigenvalue must be less than or equal to the sum of the eigenvalues, which is just the trace of $\mathbf{X}_{\perp}^T\mathbf{X}_{\perp}$. Note that $\mathbf{X}_{\perp}^T\mathbf{X}_{\perp}$ is positive semidefinite, and therefore all the eigenvalues are nonnegative (Strang, 1980). We obtain the bounds

$$R \leq \lambda_1 \leq M, \tag{11}$$

where, as previously defined, $M = \mathbf{trace}(\mathbf{X}_{\perp}^T\mathbf{X}_{\perp})$. The second equality is achieved when there is only one nonzero eigenvalue, i.e., when $\mathbf{X}_{\perp}$ is a rank 1 matrix.

The implications of Eq. (11) for fMRI experimental design are as follows. First, detection power is maximized when the columns of $\mathbf{X}_{\perp}$ are nearly parallel or, equivalently, shifted binary stimulus patterns are as similar as possible. This requirement clearly favors block designs over randomized designs in which the columns of $\mathbf{X}_{\perp}$ are nearly orthogonal. That is, the potential detection power of the block design is much greater than that of the randomized design, although as we discuss below, it is possible with some hemodynamic responses for the detection power of the block design to be less than that of a random design. Second, detection power increases with $\mathbf{trace}(\mathbf{X}_{\perp}^T\mathbf{X}_{\perp})$, which is approximately equal to the variance of the detrended binary stimulus pattern multiplied by the number of columns in $\mathbf{X}_{\perp}$. From our discussion of estimation efficiency, we know that this variance is maximized when there are an equal number of 1's and 0's in the stimulus pattern.

Although there can be some variability in the shape of the hemodynamic response, it is common to adopt an *a priori* model of the response, such as a gamma density function, when attempting to detect activations. Ideally, we would choose a design matrix for which the eigenvector $\mathbf{v}_1$ is parallel to an *a priori* response vector denoted as $\mathbf{h}_0$. With the restriction that the design matrix is constructed from binary stimulus patterns, it
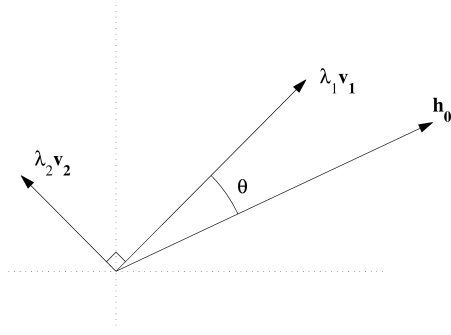
FIG. 2. Description of the angle θ between the assumed hemodynamic response $\mathbf{h}_0$ and the dominant eigenvector $\mathbf{v}_1$ of $\mathbf{X}_\perp^T\mathbf{X}_\perp$. The remaining eigenvector is denoted $\mathbf{v}_2$, and the corresponding eigenvalues are λ_1 and λ_2, respectively, where by definition $\lambda_1 \geq \lambda_2$. For an assumed $\mathbf{h}_0$, detection power is maximized when $\mathbf{v}_1$ is parallel to $\mathbf{h}_0$ ($\theta = 0$) and minimized when $\mathbf{v}_1$ is perpendicular to $\mathbf{h}_0$ ($\theta = 90°$).

may not be possible in general to achieve this goal. For each design matrix, we define θ as the angle between $\mathbf{v}_1$ and $\mathbf{h}_0$ (see Fig. 2). The achievable bound on R is then given by

$$R \leq \lambda_1 \cos^2\theta_{min} \leq M\cos^2\theta_{min}, \qquad (12)$$

where θ_{min} is the minimum angle that can be obtained over the space of all possible binary stimulus patterns. Note that θ_{min} will vary with different choices for the hemodynamic response $\mathbf{h}_0$.

On the other hand, if we have no *a priori* information about the shape of the hemodynamic response function, then a reasonable approach is to maximize the minimum value of R over the space of all possible parameter vectors $\mathbf{h}$. It is shown in the Appendix that

$$\max_{\mathbf{X}_\perp} \min_{\mathbf{h}} R \leq \frac{M}{k}, \qquad (13)$$

with equality when the columns of $\mathbf{X}_\perp$ are orthogonal and have equal energy. Therefore, in the case of no *a priori* information, the experimental design that is optimal for detection is also optimal for estimation.

Relation between Detection Power and Estimation Efficiency

We have shown that both detection power and estimation efficiency depend on the distribution of the eigenvalues of $\mathbf{X}_\perp^T\mathbf{X}_\perp$. Estimation efficiency is maximized when the eigenvalues are equally distributed, while detection power, given *a priori* assumptions about $\mathbf{h}$, is maximized when there is only one nonzero eigenvalue. In this section we explore the relation between detection power and estimation efficiency when the distribution of eigenvalues lies between these two extremes. An exception occurs in the case in which there is only one unknown parameter, i.e., $k = 1$. In

this case, there is only one eigenvalue, and the stimulus pattern that maximizes detection power is also the pattern that maximizes estimation.

We use a simple model for the distribution of eigenvalues. We assume that the maximum eigenvalue $\lambda_1 = \alpha M$ and the remaining eigenvalues are $\lambda_i = (1 - \alpha)M/(k - 1)$ where α ranges from $1/k$ to 1. This model provides a continuous transition from the case in which there is only one nonzero eigenvalue ($\alpha = 1$) to the case in which the eigenvalues are equally distributed, $\alpha = 1/k$. As the value of the dominant eigenvalue decreases, the remainder $M - \alpha M$ is equally distributed among the other eigenvalues. This equal distribution of eigenvalues results in the maximum estimation efficiency achievable for each value of the dominant eigenvalue. Assuming that the noise has unit variance, the estimation efficiency is

$$\xi(\alpha) = \frac{\alpha(1 - \alpha)M}{1 + \alpha(k^2 - 2k)}, \qquad (14)$$

which obtains a maximum value of M/k^2 at $\alpha = 1/k$. The Rayleigh quotient is

$$R(\alpha, \theta) = \left(\alpha\cos^2\theta + \frac{1 - \alpha}{k - 1}\sin^2\theta\right)M, \qquad (15)$$

where θ was previously defined. For each value of θ a parametric plot of $\xi(\alpha)$ versus $R(\alpha, \theta)$ traces out a trajectory that moves from an unequal distribution of eigenvalues at $\alpha = 1$ to an equal distribution at $\alpha = 1/k$. When the eigenvalues are equally spread, we find that $R(1/k, \theta) = M/k$, i.e., the detection power of a random design is $1/k$ times the maximum possible detection power. Note that this is also the equality relation in Eq. (13) for the detector that maximizes the minimum detection power. When $\theta = \cos^{-1}(\sqrt{1/k})$, $R(\alpha, \theta) = M\sin^2\theta/(k - 1) = M/k$ is independent of α, i.e., the plot of ξ versus R is a vertical line.

Parametric curves of $\xi(\alpha)$ versus $R(\alpha, \theta)$ for a range of dimensions k and angles θ are shown in Fig. 3. The efficiency $\xi(\alpha)$ is normalized by $\xi(1/k)$, while $R(\alpha, \theta)$ is normalized by $R(1.0, 0)$. Each curve begins at $\alpha = 1.0$ with estimation efficiency $\xi = 0$ and ends at $\alpha = 1/k$ with a normalized efficiency $\xi = 1.0$. Along the way, the curve maps out the trade-off between estimation efficiency and detection power. If $\theta < \cos^{-1}(\sqrt{1/k})$, then the detection power decreases as α decreases. However, for $\theta > \cos^{-1}(\sqrt{1/k})$, the detection power increases as α decreases, so that the random stimulus pattern with equal eigenvalues is a better detector than the initial pattern with unequal eigenvalues. It is important to emphasize here that θ depends on the assumed hemodynamic response $\mathbf{h}_0$, so that a stimulus that outperforms a random pattern for one response may perform
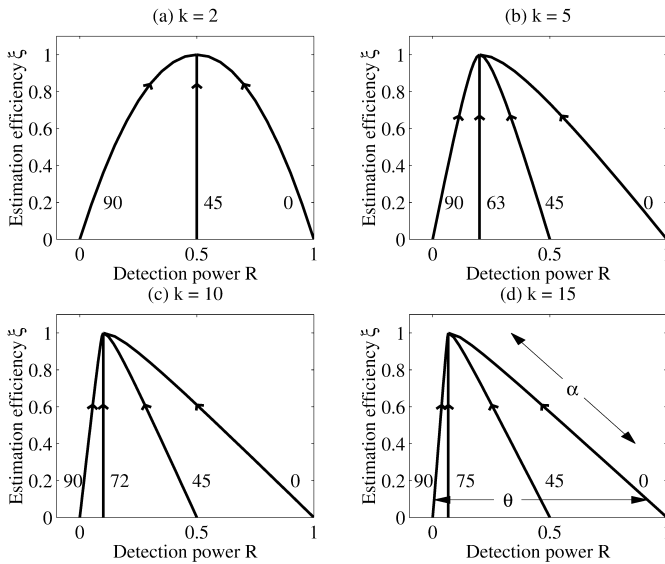
FIG. 3. Normalized estimation efficiency $\xi(\alpha)/\xi(1/k)$ versus normalized Rayleigh quotient $R(\alpha, \theta)/R(1.0, 0)$, which is a measure of detection power. Each graph corresponds to a specified dimension k of the parameter vector $\mathbf{h}$. In the parametric plots of ξ versus R, the arrows point in the direction of decreasing α, i.e., moving from $\alpha = 1$ to $\alpha = 1/k$. Each line is labeled by the angle θ between the eigenvector $\mathbf{v}_1$ and the parameter vector $\mathbf{h}$. Vertical lines correspond to $\theta = \cos^{-1}(\sqrt{1/k})$.

more poorly for another assumed response. For example, as shown under Results, a one-block design performs better than a random design when $\mathbf{h}_0$ is assumed to be a gamma density function (Fig. 5) and $\theta < \cos^{-1}(\sqrt{1/k})$. However, the one-block design performs worse than a random design when $\mathbf{h}_0$ is the first difference of the gamma density function (Fig. 8) and $\theta > \cos^{-1}(\sqrt{1/k})$.

Balancing Detection Power and Estimation Efficiency

The parametric curves defined in Eqs. (14) and (15) and plotted in Fig. 3 show that there is a fundamental trade-off between detection power and estimation efficiency. Maximum detection power comes at the price of minimum estimation efficiency, and conversely maximum estimation efficiency comes at the price of reduced detection power. The appropriate balance between power and efficiency depends on the specific goals of the experiment. At one extreme, designs that maximize detection power are optimal for experiments that aim to determine which regions of the brain are active. At the other extreme, designs that maximize estimation efficiency are optimal for experiments that aim to characterize the shape of the hemodynamic response in a prespecified region of interest. As shown in Fig. 3, there are many possible intermediate designs that lie between these two extremes. These intermediate designs may be useful for experiments in which both detection and estimation are of interest. We refer to these intermediate designs as *semirandom* designs.

In this section we present a cost criterion that can be used to select semirandom designs that achieve desired levels of estimation efficiency and detection power. The cost criterion reflects the relative time required for a design to obtain a desired level of performance. Recall that designs are parameterized by α, which reflects the relative spread of the eigenvalues. For a design with parameter α, we may determine the length of the experiment required to achieve the performance of either an optimal estimator ($\alpha = 1/k$) or an optimal detector ($\alpha = 1.0$). As an example, consider a design with a normalized estimation efficiency $\xi = 0.5$ that is half that of the optimal estimator. Since efficiency is inversely proportional to variance, we can achieve the same variance as the optimal estimator ($\xi = 1.0$) by doubling the length of our experiment. To formalize this idea we define a relative estimation time,

$\tau_{est}(\alpha) =$ relative time to achieve desired efficiency

$$= \frac{(\text{maximum possible efficiency}) \times f_{est}}{\text{efficiency of this design}},$$

where f_{est} is the fraction of the maximum possible estimation efficiency that we want to achieve. For example $f_{est} = 0.75$ corresponds to an experiment in which we want to obtain 75% of the efficiency of an optimal estimator. If the normalized efficiency of the design is $\xi = 0.5$, then the relative estimation time is $\tau_{est}(\alpha) = 0.75 \times 1.0/0.5 = 1.5$. This means that we would need to increase the length of an experiment with $\xi = 0.5$ by 50% in order to achieve 75% of the maximum possible efficiency. In a similar fashion we define the relative detection time as

$\tau_{\det}(\alpha, \theta) =$ relative time to achieve desired power

$$= \frac{(\text{maximum possible detection power}) \times f_{\det}}{\text{detection power of this design}},$$

where $f_{\det}$ is the fraction of the maximum possible detection power that we want to achieve. Assuming that the desired detector has greater detection power than a random design (i.e., $\theta < \cos^{-1}(\sqrt{1/k})$), the relative detection power $\tau_{\det}(\alpha, \theta)$ decreases monotonically with α, since the maximum detection power is obtained when there is only one nonzero eigenvalue. On the other hand, we find that the relative estimation time $\tau_{est}(\alpha)$ increases monotonically with α, since estimation efficiency decreases as the eigenvalues become more unequally distributed.

For each value of α, the time required to obtain both the desired efficiency and the desired power is

$$\tau(\alpha, \theta) = \max[\tau_{est}(\alpha), \tau_{\det}(\alpha, \theta)],$$

i.e., the greater of the relative estimation time and the detection time. We argue that the best design is the one that minimizes $\tau(\alpha, \theta)$. Because $\tau_{est}(\alpha)$ increases with α and $\tau_{\det}(\alpha, \theta)$ decreases with α, a unique minimum occurs at $\tau_{est}(\alpha) = \tau_{\det}(\alpha, \theta)$, the point at which the relative times intersect. We refer to the value of the minimum as τ_{opt} and the optimal value of α as α_{opt}. Analytical expressions for $\tau_{est}(\alpha)$, $\tau_{\det}(\alpha, \theta)$, τ_{opt}, and α_{opt} are provided in the Appendix.

As an example of a semirandom design that satisfies the minimum time criterion, we first examine the case in which $k = 15$, $\theta = 45°$, $f_{det} = 1.0$, and $f_{est} = 1.0$. From the equations in the Appendix, the minimum-time design occurs for $\alpha_{opt} = 0.52$ and $\tau_{opt} = 1.8$. This design simultaneously achieves maximum estimation efficiency and detection power at the cost of an 80% increase in experimental time. It lies roughly halfway between a random design (orthogonal) and a block design (highly nonorthogonal).

We next consider an example in which the cost criterion can aid in the generation of a new type of design that we refer to as a *mixed* design. This design is the concatenation of a block design and a semirandom design. We begin with a one-block design of length N, which for the purpose of this example we assume to have a normalized detection power of 1.0 and a normalized estimation efficiency of 0.0. A shorter one-block design of length rN that has the same fraction of 1's as in the original design will have a normalized detection power r. If we concatenate this shorter block design with a semirandom design, the detection power of the semirandom design should be $(1 - r)$ in order for the mixed design to have a detection power of 1.0. Also, the efficiency of the semirandom design should be 1.0, since the block design has an estimation efficiency of 0. The semirandom design that satisfies these requirements can be found from the equations in the Appendix with $f_{det} = 1 - r$ and $f_{est} = 1.0$. The length of the semirandom design is τ_{opt} and the design is characterized by the parameter α_{opt}.

Figure 4 shows two examples of mixed designs and one example of a semirandom design. The uppermost design consists of a one-block design with relative length $r = 0.8$ concatenated with a random design with relative length $\tau_{opt} = 1.0$ and design parameter $\alpha_{opt} = 1/k = 0.07$. The second mixed design consists of a one-block design with reduced length $r = 0.5$ concatenated with a semirandom design with length $\tau_{opt} = 1.3$ and design parameter $\alpha_{opt} = 0.33$. Finally, the lowermost design is a semirandom design with $\tau_{opt} = 1.8$ and design parameter $\alpha_{opt} = 0.51$. Note that the total relative length of each of the designs is 1.8. In addition, although the three designs look very different, the estimation efficiency and detection power across the three designs are identical. In order to achieve this property, the semirandom design becomes increasingly
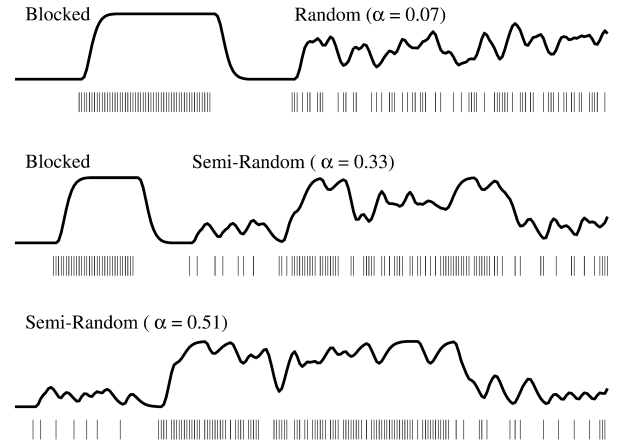

FIG. 4. Mixed and semirandom design examples. The estimation efficiency and detection power are identical across designs. The uppermost design consists of a one-block design followed by a random design. The middle design consists of a shorter one-block design followed by a semirandom design that has greater detection power than the random design. The lowermost design is a semirandom design that simultaneously achieves maximum estimation efficiency and detection power (i.e., $f_{det} = 1.0$, $f_{est} = 1.0$) at the cost of increased experimental length.

more block-like (e.g., increasing values of α) as the length of the block design is reduced.

Perceived Randomness of a Pattern

In the previous section, we considered the trade-off between estimation efficiency and detection power and presented a metric for the relative temporal cost of each trade-off point. While it is important to understand this trade-off, there is an additional factor that must also be considered in some fMRI experiments. This is the perceived randomness of a sequence. Randomness in a design may be critical for circumventing experimental confounds such as habituation and anticipation (Rosen *et al.,* 1998). A semirandom or mixed design that is optimal from the point of view of estimation efficiency and detection power may not provide enough randomness for a given experiment. While it is beyond the scope of this paper to address the question of how much randomness is sufficient, it is useful to define a metric for randomness so as to better understand the relationship between randomness, estimation efficiency, and detection power.

As one possible metric for perceived randomness, we consider the average "predictability" of a sequence, defined as the probability of a subject correctly guessing the next event in the sequence. A random sequence has an average predictability of 0.5, while a deterministic sequence such as a block design has an average predictability approaching 1.0. As described under Methods, the predictability can be gauged either with a computer program or by

measuring how well a population of human subjects can predict a given sequence.

METHODS

We calculated estimation efficiencies and detection powers using a linear model with $k = 15$ and $N = 128$. The dimension of the interference subspace was varied from 1 to 4, with Legendre polynomials of order 0 to 3 forming the columns of the matrix **S**. Semirandom stimulus patterns with $m = 64$ were obtained by permuting various block designs (Buxton *et al.,* 2000). We used block designs with 1 to 32 equally sized and spaced blocks and at each permutation step exchanged the positions of two randomly chosen events. The relative shift of each block design was chosen to make the pattern as orthogonal as possible to the interference subspace—this shift is in general dependent on the dimension of the interference subspace. A total of 80 permutation steps were performed for each block design, and the estimation efficiency and detection power were computed at each step. In addition, 1000 patterns with a uniform distribution of 1's in the pattern were generated, and the 30 patterns with the greatest estimation efficiency were used for further analysis. For calculation of detection power, the parameter vector **h** was a gamma density function of the form $h[j] = (\tau n!)^{-1}$ $(j\Delta t/\tau)^{n}e^{-j\Delta t/\tau}$ for $j \geq 0$ and 0 otherwise (Boynton *et al.,* 1996). We used gamma density functions with τ ranging from 0.8 to 1.6 and n taking on values of either 2 or 3. In all cases, we used $\Delta t = 1$. Examples of these gamma density functions are shown in Fig. 7. We also calculated the detection power with a parameter vector that is the first difference of the gamma density function. As shown in Fig. 8, this vector exhibits an initial increase followed by a prolonged undershoot. The area of the vector is essentially zero, and the frequency response is bandpass, meaning that it is zero at zero frequency, increases with frequency, attains a maximum at some peak frequency, and then decreases with frequency.

To measure the average predictability of each pattern, we used a binary string prediction program based on the work of Fudenberg and Levine (1999) to predict the events in each stimulus pattern (code can be obtained from http://levine.sscnet.ucla.edu//Games/binlearn.htm). This program uses a lookup table of past events to generate conditional probabilities for the next event. In preliminary tests, the scores generated by the program were found to be in good agreement with scores generated by three human volunteers.

RESULTS

Figure 5 shows the paths of estimation efficiency versus detection power for the random designs and the various permuted block designs. The parameters for **h**
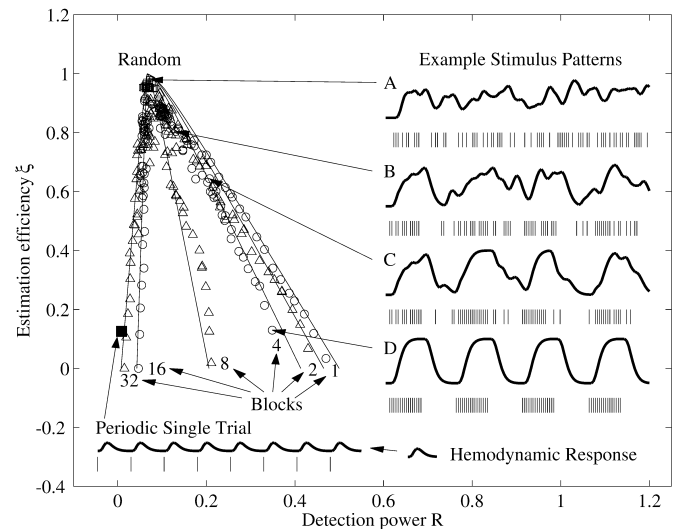

FIG. 5. Simulation results for estimation efficiency versus detection power in which the interference subspace is limited to a constant term and the hemodynamic response parameters are $\tau = 1.2$ and $n = 3$. Paths of open symbols are labeled by the number of blocks in the original block design and show the performance as the block design is randomly permuted. For all designs $m = 64$ and $N = 128$. Theoretical curves (solid lines) are also shown, with the angles corresponding to 1, 2, 4, 8, 16, and 32 blocks set equal to 45, 47, 50, 63, 78, and 85°, respectively. Example stimuli and responses based on permutations of the 4-block design are shown on the right-hand side. A is a random design, B and C are semirandom, and D is the block design. The performance and stimulus pattern for a periodic single-trial experiment are shown in the lower left-hand corner.

were $\tau = 1.2$ and $n = 3$, corresponding to response II in Fig. 7; these are also the response parameters used in Figs. 6, 8, and 9. The dimension of the interference subspace was $l = 1$, meaning that only a constant term was removed from the columns of the design matrix. The paths taken by the permuted designs are well-modeled by theoretical curves. This reflects the fact that as the block design becomes increasingly randomized, the distribution of eigenvalues of $\mathbf{X}_{\perp}^{T}\mathbf{X}_{\perp}$ becomes more even. Note that the permutation algorithm does not explicitly try to equalize the spread of eigenvalues, so that in some cases the path taken by the permuted design can deviate significantly from the theoretical curve, e.g., the path for eight blocks. In addition, it is important to note that we have shown only one realization of the permuted paths—since the permutation procedure is random, many paths are possible, and some will follow the theoretical curves better than others. Upon examination of many realizations, we have found that the theoretical curves capture the overall behavior of the permuted patterns as they migrate toward a random design.

The 1-block design has the greatest detection power for the assumed gamma density function parameter vector. The angle between the parameter vector **h** and the dominant eigenvector of $\mathbf{X}_{\perp}^{T}\mathbf{X}_{\perp}$ for this design is about 45°, so that its detection power is half that of a
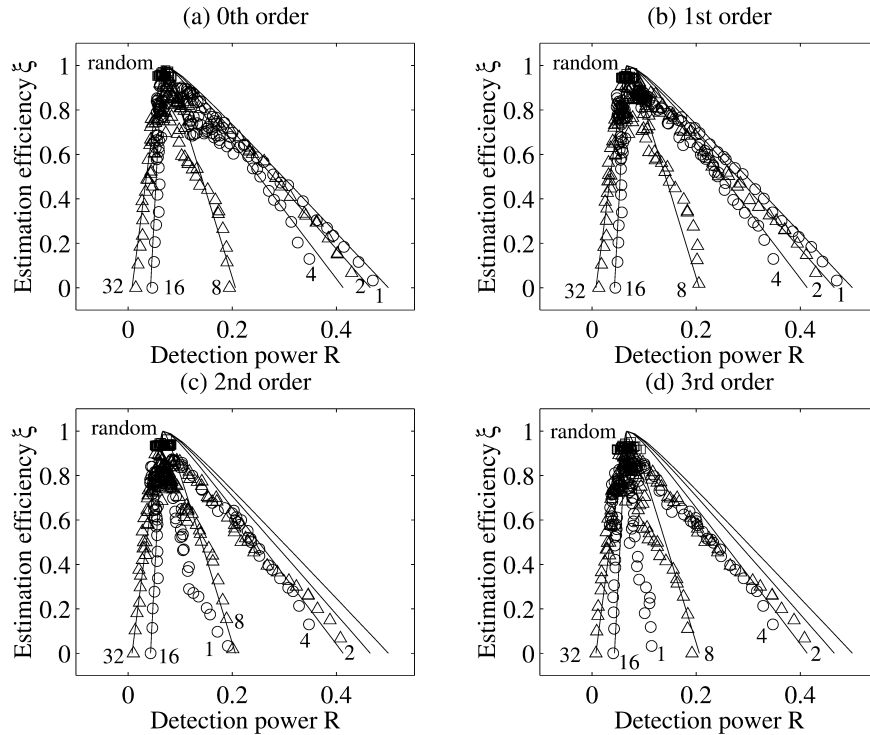
FIG. 6. Estimation power versus detection power with removal of nuisance effects. Each plot is labeled by the highest order of Legendre polynomial that is included in the interference subspace model. Paths of open symbols are labeled by the number of blocks in the design prior to permutation. Theoretical curves use the angles listed in Fig. 5. Other parameters: $m = 64$, $N = 128$.

design in which the dominant eigenvector is parallel to **h**. It is not clear if it is possible to achieve a smaller angle using binary stimulus patterns. The 32-block design has the smallest detection power because its stimulus pattern has the highest fundamental frequency and the magnitude response of the gamma density function falls off with frequency.

Example stimulus patterns and responses (stimulus pattern convolved with **h**) for four points along the path for the permuted 4-block design are shown on the right-hand side of Fig. 5. Stimulus pattern A corresponds to a random design, B and C are semirandom designs, and D is the block design. The semirandom designs retain the overall shape of the block design with enough randomness added in to obtain significant increases in estimation efficiency.

The performance of a periodic single trial design with one trial every 16 s is shown in the lower left-hand corner of Fig. 5. Both the estimation efficiency and the detection power are low because the number of events is only $m = 8$, which is much smaller than the number of events, $N/2 = 64$, that maximizes both efficiency and power. As a consequence the trace of $\mathbf{X}_\perp^T \mathbf{X}_\perp$ is much smaller than the bound stated in Eq. (5).

Figure 6 shows the estimation efficiency and detection power for the permuted paths as the dimension of the interference subspace is increased from 1 to 4. When the dimension of the subspace is $l = 4$, the

projection operator $\mathbf{P}_S^\perp$ removes a constant term, a linear term, a quadratic term, and a cubic term from the columns of **X**. The detection power of the 1-block design
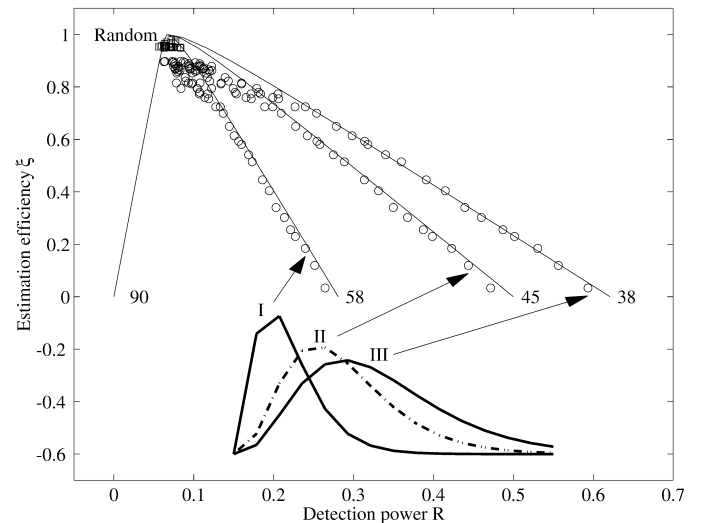

FIG. 7. Estimation efficiency and detection power with permuted versions of the one-block design and three different hemodynamic responses. The parameters for the hemodynamic responses are I, $\tau = 0.8$, $n = 2$; II, $\tau = 1.2$, $n = 3$; and III, $\tau = 1.6$, $n = 3$. The responses are normalized to have equal energies. The area, and hence the low-frequency gain, of response I is smaller than that of response II, which is in turn smaller than that of response III. Theoretical curves are labeled by the value of θ.
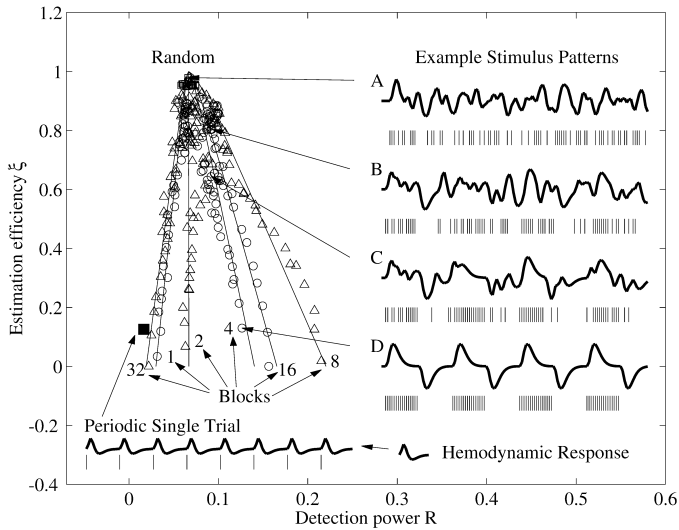
FIG. 8. Simulation results for estimation efficiency versus detection power when the hemodynamic response is the first difference of the hemodynamic response used in Fig. 5. Paths of open symbols are labeled by the number of blocks in the original block design and show the performance as the block design is randomly permuted. The permutations are identical to those used in Fig. 5. Theoretical curves (solid lines) are also shown, with the angles corresponding to 1, 2, 4, 8, 16, and 32 blocks set equal to 80, 75, 68, 62, 66, and 82°, respectively. Example stimuli and responses based on permutations of the 4-block design are shown on the right-hand side. A is a random design, B and C are semirandom, and D is the block design. The performance and stimulus pattern for a periodic single-trial experiment are shown in the lower left-hand corner. Note that the horizontal scale is about half that of Fig. 5.

is severely reduced after removal of quadratic and cubic terms, while the detection power of the 2-block design is less affected. Random designs and block designs starting with four or more blocks are relatively unaffected. The sensitivity of the 1-block design results from the fact that its low frequency content is greater than designs with more blocks (Frackowiak *et al.*, 1997). The 4-block design offers robustness to nuisance terms while maintaining good detection power. The angle between the dominant eigenvector and the parameter vector **h** for this design is about 50°, and its detection power is roughly 80% of the maximum detection power of the 1-block design (i.e., the detection power of the 1-block design when the interference consists of a constant term only).

In Figs. 7 and 8, we consider the variability of the hemodynamic response function. From the theory section, we know that estimation efficiency does not depend on the parameter vector **h**. As a result, variations in **h** affect only the detection power of a design. We may also view this as a process in which varying **h** simply changes the angle θ between **h** and the dominant eigenvector of a design.

Figure 7 shows the estimation efficiency and detection power for permuted versions of the 1-block design for three different hemodynamic response functions,

ranging from a narrow response with τ = 0.8 and n = 2 to a broad response with τ = 1.6 and n = 3. As the hemodynamic response broadens, we find that the detection power increases or, equivalently, the angle θ decreases. This is because the dominant eigenvector of the 1-block design is rather broad. These changes in detection power are further examined under Discussion.

Figure 8 shows the estimation efficiency and detection power assuming a parameter vector **h** that is the first difference of the gamma density function used in Fig. 5. The estimation efficiencies are identical to those shown in Fig. 5, but the detection powers are significantly different. Whereas we previously found that detection power decreased with the number of blocks, we now find that the detection power increases with the fundamental frequency of the stimulus pattern (i.e., moving from a 1-block design to an 8-block design), attains a maximum with the 8-block design, whose fundamental frequency is closest to the peak frequency of the bandpass response of **h**, and then decreases as the fundamental frequency exceeds the peak frequency (i.e., moving from an 8-block to a 32-block design). These changes in detection power are well described by adjusting the angle θ in the theoretical model.

Figure 9 shows contours of average predictability, as computed using the binary string prediction computer program, superimposed on a grid of normalized estimation efficiency versus detection power. The permutation path for the four-block design is also shown. Irregularities in the contours are due to the fact that the estimation efficiency and detection power of the permuted block designs do not follow smooth trajectories.
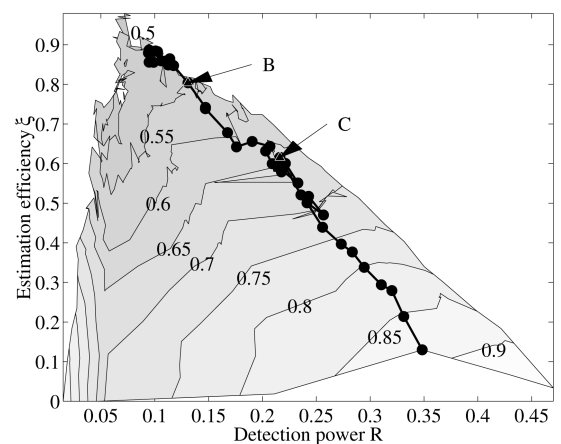

FIG. 9. Estimation efficiency, detection power, and predictability. Contours are labeled by predictability index. The connected solid dots show the permutation path for the four-block design. The predictabilities of points B and C are 0.55 and 0.63, respectively. The detection power of point B is approximately twice that of random designs, which have a predictability of 0.5. The stimulus patterns and responses for B and C are shown in Fig. 5.

Average predictability decreases as estimation efficiency increases, with random patterns having an average predictability of about 0.5. The semirandom pattern that meets the minimum-time criterion (with $f_{est} = f_{det} = 1.0$) for permuted 4-block designs has a predictability of 0.63 (point C in Fig. 9). To obtain a lower predictability of 0.55, it is necessary to select a design (point B) that has 30% higher normalized estimation efficiency (0.80 vs 0.61) and 40% lower normalized detection power (0.13 vs 0.22), compared to the minimum-time design. The relative times required to achieve the detection power of the block design (with detection power 0.41) are 3.2 and 1.9 for points B and C, respectively. Although the relative time for point B is 65% higher than that of the minimum-time point (C), it is only half the relative time for a random design, which has a normalized detection power of 0.066 (i.e., $1/k$ where $k = 15$).

DISCUSSION

Detection power, estimation efficiency, and predictability represent three key features of any experimental design for fMRI. The choice of design is critical, for it is possible to select a design that performs poorly in all three respects, e.g., periodic single-trial designs. The work in this paper provides a theoretical framework for understanding the bounds on and the relationship between estimation efficiency and detection power. We believe that this framework will be useful for assessing the relative merits of proposed designs and in determining if better designs are possible. At this point, we lack a theory that relates predictability to efficiency and power. However, simulations such as those shown in Fig. 9 can be used to understand the relative trade-offs.

A key aspect of our work is the demonstration that the relation between efficiency and power is characterized by two parameters: α and θ. The parameter α describes the relative spread of the eigenvalues associated with the design and to first order reflects the randomness of a design, with $\alpha = 1/k$ corresponding to a random design and $\alpha = 1.0$ corresponding to a nonrandom design such as a block design. The parameter θ is the angle between the parameter vector **h** and the dominant eigenvector of the design and to first order reflects how close a nonrandom design ($\alpha = 1.0$) is to achieving the maximum possible detection power, with the maximum achieved when $\theta = 0$. By varying α and θ we can easily map the trade-off between efficiency and power without making any assumptions about the specifics of the design or the parameter vector **h**. As a result, the theoretical trade-off curves provide bounds on the performance of all possible designs. In addition they serve as a framework for understanding the performance of specific designs such as those shown in Figs. 5 and 8.

To place this work in context, we note that the importance of estimation efficiency in the context of fMRI was introduced by Dale (1999). The use of the F statistic in fMRI is due to Friston and co-workers (Friston *et al.*, 1995b), as is the use of dynamic stochastic designs (analogous to the semirandom designs in this paper) as an intermediate trade-off between random and block designs (Friston *et al.*, 1999). In Friston *et al.* (1999), the emphasis is placed on the $k = 1$ case in which maximizing estimation efficiency is equivalent to maximizing detection power. Our work extends that of Dale (1999) and Friston *et al.* (1999) by considering the relation between efficiency and power when k is greater than 1.

There are a number of interesting issues that are beyond the scope of this paper. We discuss these in the following paragraphs.

How Much Randomness?

We have proposed the average predictability of a sequence as a measure of perceived randomness, but further work is required to determine how unpredictable a sequence needs to be in order to sufficiently minimize psychological confounds. It is likely that the correct answer will depend on the specifics of the experiment at hand. The results presented in Fig. 9 show that a semirandom pattern that is slightly more predictable (e.g., predictability = 0.55) than a random pattern yields a 100% increase in detection power and only a 20% decrease in estimation efficiency, with respect to a random design. Thus, if slight increases in predictability are acceptable for a given experimental paradigm, the advantages from the point of view of statistical efficiency can be significant.

Generation of Stimulus Patterns

With the framework developed in this paper, we can assess the relative merits of various experimental designs. However, the generation of optimal patterns is an open problem. Our method of randomly permuting block designs is promising, but it is not guaranteed to find the optimal pattern. It is possible that numerical optimization methods may be more efficient for finding stimulus patterns with a desired distribution of eigenvalues or a desired ratio of estimation efficiency to detection power. The framework of dynamic stochastic designs proposed by Friston *et al.* (1999) may also offer a method of reducing the time required to search for optimal patterns.

Variability of Hemodynamic Responses

Hemodynamic responses exhibit a wide variability in shapes, especially across subjects and possibly across cortical areas (Aguirre *et al.*, 1998). We have shown that the effect of this variability on detection power can

be described by the angle between the response vector and the dominant eigenvector of the design. However, as we discuss below, there is a subtle point that needs to be understood with respect to these changes in detection power.

In interpreting Fig. 7, it is important to note that the plot does not necessarily imply that wider hemodynamic responses are more detectable than narrower responses. In the definition of the Rayleigh quotient, the noncentrality parameter is normalized by the energy of the hemodynamic response. As shown under Theory, the energy normalization is convenient for understanding the dependence of detection power on the structure of the design matrix. In addition, it provides a measure of how close a design is to achieving the absolute maximum detection power for a given hemodynamic response. For example, Fig. 7 shows that a one-block design is closer to achieving the maximum possible detection power for a wider response versus a narrow response.

Energy normalization allows us to compare the detection power across hemodynamic shapes with the same energies. However, other normalizations may be more instructive. For example, using the area under the hemodynamic response as a normalization factor would be consistent with a picture in which neural activity gives rise to a fixed increase in blood volume that is then delivered over a time interval that varies from subject to subject. With area normalization, we would find that the detection powers for the various response shapes shown in Fig. 7 are approximately the same due to the fact that the spectral amplitudes of the responses at low frequencies are directly proportional to their area. Area normalization would not be as useful, though, in comparing the detection power of a response that is the first difference of the gamma density function (Fig. 8) and has zero area to that of a gamma density function response, which has nonzero area. A more meaningful normalization in this case would be the peak spectral magnitude of the response.

Finally, the detection power with any choice of normalization can be simply related to the Rayleigh quotient. In the case of area normalization, the detection power normalized by area is $\mathbf{h}^T\mathbf{h}/\mathbf{1}^T\mathbf{h}\ R$, where $\mathbf{1}$ is a column vector of 1's.

Multiple Event Types

There is increasing interest in fMRI experiments in which the responses to multiple event types are compared and contrasted (Friston *et al.,* 1998a). An extension of the theoretical framework presented here should be useful in clarifying the trade-offs between estimation efficiency and detection power for multiple event types.

Correlated Noise

In this paper, we have assumed that the additive noise term $\mathbf{n}$ in the general linear model is uncorrelated noise with covariance matrix $\mathbf{C_n} = \sigma^2\mathbf{I}$. It is straightforward to modify the definitions of estimation efficiency (Dale, 1999) and detection power to accommodate the more general case in which the covariance matrix is not a multiple of identity. However, the variability in the structure of the covariance matrix across subjects and experimental conditions complicates the selection of an optimal design prior to the experiment. One possibility is to assume a simple form, such as a first-order autoregressive model, for the covariance matrix (Dale, 1999). The impact of correlations in the noise on the relation between estimation efficiency and detection power is a subject for future work. Methods for removal of physiological noise (e.g., Glover *et al.,* 2000) may be helpful for reducing correlations in the noise to a level at which they may be safely neglected.

Nonlinearities

We have assumed that the neuronal and hemodynamic pathway from the stimulus to the measured response is well modeled as a linear time-invariant system, so that the measured response is the convolution of the stimulus with a hemodynamic response function. While the linear time-invariant approximation works reasonably well (Boynton *et al.,* 1996; Dale and Buckner, 1997), there is growing evidence that a nonlinear, time-varying model more accurately describes the pathway (Boynton *et al.,* 1996; Buxton *et al.,* 1998; Friston *et al.,* 1998b, 1999; Miller *et al.,* 2000). For example, Bandettini and Cox (2000) have shown that the measured detection power of periodic single-trial designs is higher than would be predicted by a linear time-invariant model. An extension of the theoretical framework presented in this paper to address nonlinear and time-varying effects would be of great interest.

CONCLUSION

There is a fundamental trade-off between estimation efficiency and detection power in experimental designs for fMRI. We have presented a theoretical framework that describes this trade-off and provides insight into the performance of random and block designs, as well as novel designs such as semirandom and mixed designs. We also introduced predictability as an important third factor that should be considered along with detection power and estimation efficiency in the design of an experiment. Small increases in the predictability of a sequence can yield significant gains in detection power with a minimal reduction of estimation efficiency.

APPENDIX

Approximate Bound on the Trace Term M

Assuming a uniform distribution of m 1's out of N points in the stimulus pattern, the expected number of 1's in the qth column of $\mathbf{X}$ is $m(1 - (q - 1)/N)$. The expected energy in the qth column of $\mathbf{X}_\perp$, with the mean removed, is given by $(1 - (1 - (q - 1)/N)m/N)(1 - (q - 1)/N)m$. The trace term is the sum of the energies of the columns of $\mathbf{X}_\perp$ and is approximately bounded above as $M \leq \sum_{q=1}^{k}(1 - (1 - (q - 1)/N)m/N)(1 - (q - 1)/N)m$. This bound is used in the plots shown under Results.

Proof That Orthogonal Designs Maximize the Minimum Detection Power

We want to show that $\max_{\mathbf{X}_\perp} \min_{\mathbf{h}} R \leq M/k$, with equality when the columns of $\mathbf{X}_\perp$ are orthogonal and have equal energy. The minimum of the Rayleigh quotient is equal to the minimum eigenvalue of $\mathbf{X}_\perp^T \mathbf{X}_\perp$ (Strang, 1980). The minimum eigenvalue is maximized when all of the eigenvalues are equal. From the eigenvector decomposition of $\mathbf{X}_\perp^T \mathbf{X}_\perp$, we have $\mathbf{X}_\perp^T \mathbf{X}_\perp = \mathbf{V} \Lambda \mathbf{V}^{-1} = M/k \, \mathbf{I}$, where $\mathbf{V}$ is the matrix of eigenvectors and Λ is the diagonal matrix of eigenvalues. Thus, the columns of $\mathbf{X}_\perp$ are orthogonal and have equal energy.

Expressions for Balancing Estimation Efficiency and Detection Power

The relative time to achieve the desired efficiency is

$$\tau_{est}(\alpha) = f_{est} \frac{\xi(\alpha = 1/k)}{\xi(\alpha)} = \frac{f_{est}}{k^2} \frac{1 + \alpha(k^2 - 2k)}{\alpha(1 - \alpha)}.$$

The relative time to achieve the desired detection power is

$$\tau_{det}(\alpha, \theta) = f_{det} \frac{R(\alpha = 1.0, \theta)}{R(\alpha, \theta)} = \frac{f_{det}\cos^2\theta}{\alpha \cos^2\theta + \dfrac{1 - \alpha}{k - 1} \sin^2\theta}.$$

The point at which $\tau_{est}(\alpha) = \tau_{det}(\alpha, \theta)$ is

$$\alpha_{opt} = \frac{-b + \sqrt{b^2 - 4ac}}{2a},$$

where

$$a = (k^2 - 2k)\left(\cos^2\theta - \frac{\sin^2\theta}{k - 1}\right) + k^2 \frac{f_{det}}{f_{est}} \cos^2\theta,$$

$$b = (k^2 - 2k - 1)\frac{\sin^2\theta}{k - 1} + \left(1 - \frac{f_{det}}{f_{est}} k^2\right)\cos^2\theta,$$

$$c = \frac{\sin^2\theta}{k - 1}.$$

τ_{opt} is then obtained by inserting α_{opt} into the expression for $\tau_{est}(\alpha)$.

REFERENCES

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

Bandettini, P. A., and Cox, R. W. 2000. Event-related fMRI contrast when using constant interstimulus interval: Theory and experiment. *Magn. Reson. Med.* **43:** 540–548.

Behrens, R. T., and Scharf, L. L. 1994. Signal processing applications of oblique projection operators. *IEEE Trans. Signal Processing* **42:** 1413–1424.

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

Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Peterson, S. E., Raichle, M. E., and Rosen, B. R. 1996. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA* **93:** 14878–14883.

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

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

Buxton, R. B., Liu, T. T., Martinez, A., Frank, L. R., Luh, W.-M., and Wong, E. C. 2000. Sorting out event-related paradigms in fMRI: The distinction between detecting an activation and estimating the hemodynamic response. *NeuroImage* **11:** S457.

Buxton, R. B., Wong, E. C., and Frank, L. R. 1998. Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magn. Reson. Med.* **39:** 855–864.

Dale, A. M. 1999. Optimal experimental design for event-related fMRI. *Hum. Brain Mapp.* **8:** 109–114.

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

Draper, N. R., and Smith, H. 1981. *Applied Regression Analysis.* Wiley, New York.

Frackowiak, R. S. J., Friston, K. J., Frith, C. D., Dolan, R. J., and Mazziotta, J. C. 1997. *Human Brain Function.* Academic Press, San Diego.

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

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., and Frackowiak, R. S. J. 1995b. 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., and Turner, R. 1998a. Event-related fMRI: Characterizing differential responses. *NeuroImage* **7:** 30–40.

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

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N. A., and Dale, A. M. 1999. Stochastic designs in event-related fMRI. *NeuroImage* **10:** 607–619.

Fudenberg, D., and Levine, D. K. 1999. Conditional universal consistency. *Games Econ. Behav.* **29:** 104–130.

Glover, G. H., Li, T.-Q., and Ress, D. 2000. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn. Reson. Med.* **44:** 162–167.

Josephs, O., and Henson, R. N. A. 1999. Event-related functional magnetic resonance imaging: Modelling, inference and optimization. *Philos. Trans. R. Soc. London B* **354:** 1215–1228.

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

Miller, K. L., Luh, W.-M., Liu, T. T., Martinez, A., Obata, T., Wong, E. C., Frank, L. R., and Buxton, R. B. 2000. Characterizing the dynamic perfusion response to stimuli of short duration. In *Proceedings of the Eighth Meeting, International Society for Magnetic Resonance in Medicine, Denver, Colorado,* p. 500.

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

Scharf, L. L., and Friedlander, B. 1994. Matched subspace detectors. *IEEE Trans. Signal Processing* **42:** 2146–2157.

Seber, G. A. F. 1977. *Linear Regression Analysis.* Wiley, New York.

Strang, G. 1980. *Linear Algebra and Its Applications.* Harcourt Brace Jovanovich, San Diego.

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