# Optimization of experimental design in fMRI: a general framework using a genetic algorithm

Tor D. Wager* and Thomas E. Nichols

*Department of Psychology, C/P Area, University of Michigan, 525 E. University, Ann Arbor, MI 48109-1109, USA*

## Abstract

This article describes a method for selecting design parameters and a particular sequence of events in fMRI so as to maximize statistical power and psychological validity. Our approach uses a genetic algorithm (GA), a class of flexible search algorithms that optimize designs with respect to single or multiple measures of fitness. Two strengths of the GA framework are that (1) it operates with any sort of model, allowing for very specific parameterization of experimental conditions, including nonstandard trial types and experimentally observed scanner autocorrelation, and (2) it is flexible with respect to fitness criteria, allowing optimization over known or novel fitness measures. We describe how genetic algorithms may be applied to experimental design for fMRI, and we use the framework to explore the space of possible fMRI design parameters, with the goal of providing information about optimal design choices for several types of designs. In our simulations, we considered three fitness measures: contrast estimation efficiency, hemodynamic response estimation efficiency, and design counterbalancing. Although there are inherent trade-offs between these three fitness measures, GA optimization can produce designs that outperform random designs on all three criteria simultaneously.
© 2003 Elsevier Science (USA). All rights reserved.

## Introduction

This article describes a method for optimizing the stimulus sequence in event-related fMRI using a genetic algorithm (GA). We describe how genetic algorithms may be applied to experimental design for fMRI, and we present simulations of random and optimized stimulus sequences over a range of design parameters. Event-related fMRI is a major advance in methodology because of the flexibility it allows in experimental design and the specificity with which particular mental processes may be studied; but the space of possible design choices for a given study is large and complex, and it is difficult to know which parameter values are best. The framework for optimization we present here is general and can be used to maximize statistical and non-statistical (i.e., psychological validity) properties of fMRI designs using novel criteria or existing criteria for design fitness (e.g., Dale, 1999; Friston et al., 1999, 2000a; Josephs and Henson, 1999; Liu et al., 2001). Previous research on optimization has considered only a narrow range of possible parameter values and design types, often excluding factors such as temporal autocorrelation of fMRI noise, nonlinearity in observed signal, the presence of multiple conditions and multiple contrasts of interest within a single experiment, experiment-related factors such as psychological probes that influence the design but are difficult to model, and factors such as counterbalancing of stimuli and repeated presentations that influence the psychological validity of the task. The flexibility of the genetic algorithm as an optimization tool, combined with novel methods for estimating signal nonlinearities, allows us to circumvent all of these limitations. Within the GA framework, researchers can develop experimental designs that are optimal with respect to the unique characteristics of the research paradigm under investigation.

Optimization of experimental designs may best be described as a search through the space of possible designs,

* Corresponding author.

*E-mail address:* torw@umich.edu (T.D. Wager).

with the dimensionality of the space defined by the number of design parameters allowed to vary. We combine use of the genetic algorithm with conventional exhaustive search over subsets of design parameters. A conventional search is used when the search space is very restricted and exhaustive search is possible. For example, when choosing an inter-stimulus interval (ISI), the search proceeds over a single-parameter dimension, and design fitness is assessed over a range of ISIs. This type of parameter mapping is feasible over a low-dimension parameter space (i.e., when trying to choose one or two design parameters). However, as Dale (1999) discovered, there is great variability among realizations of random event-related designs with the same parameter values. Optimization of the particular ordering of events is accomplished with the genetic algorithm, which is effective at optimizing over very high-dimensional parameter spaces, and yields a particular pseudorandom sequence of events with maximum suitability.

Optimization of event sequences promises the greatest improvement in experimental fMRI design when rapid, event-related sequences are used. The advantages of rapid event-related fMRI over blocked fMRI or PET designs have been discussed at length (Buckner, 1998; Burock et al., 1998; Dale & Buckner, 1997; D'Esposito et al., 1999; Friston et al., 1998; Josephs and Henson, 1999; Rosen et al., 1998). Briefly, one major advantage of event-related designs is that they permit more specific isolation of particular psychological events than block designs, making psychological inference easier and making it possible to image a whole new class of psychological events. Statistically, rapid designs (typically with ISIs < 4 s) may improve statistical power by as much as 10:1 over single-trial designs (Dale, 1999). Not only is it possible to estimate the magnitude of regional activation using rapid designs, it is also possible to obtain reliable estimates of the time course of activation with subsecond resolution (Miezin et al., 2000).

However, all these advantages are mitigated by serious limitations posed by the shape and linearity of the signal response (e.g., BOLD) in fMRI. The statistical power of effects in rapid event-related fMRI depends greatly on the timing parameters, the particular sequence of events chosen, and how these variables interact with signal nonlinearities, temporal noise autocorrelation, and psychological factors relating to the consistency and magnitude of the expected neural activity across trials.

In this article, we first review previous approaches to experimental design optimization in fMRI. We then introduce the genetic algorithm and the way in which it is applied in this context. Under Methods, we describe the linear systems approach and a modification to account for nonlinearity. Following that, we describe the measures of design efficiency and the method for parameterizing an fMRI model. We then describe measures of psychological suitability (e.g., counterbalancing). Finally, we describe four simulations demonstrating the ability of the genetic algorithm to find optimal designs for various combinations of statistical and psychological considerations.

### Previous approaches to optimization of fMRI designs

Josephs and Henson (1999) were among the first to examine the relative fitness of fMRI designs in a linear system. They generated random event sequences with specified transitional probabilities, performed an exhaustive search across various ISIs, and examined designs with and without temporal "jitter" or stochastic variation in event onset times (Burock et al., 1998; Dale, 1999). Their measure of fitness was "estimated measurable power" (EMP)—essentially the average squared deviation of a predictor from its mean.

The use of EMP as a criterion of fitness is a reasonable choice, as the statistical power of the design increases monotonically as the energy (EMP) of the predictor increases. However, EMP is limited because it does not take into account the other major constraint on statistical power: correlation between predictors in the design matrix.

Dale (1999) provided an alternate measure of design fitness, termed efficiency. Efficiency is the inverse of the sum of the variance of parameter estimates (see Methods, Eq. (7)). Increased variance of parameter estimates reduces the significance for a given effect size, so maximizing efficiency is a reasonable endeavor. In fact, in the discipline of experimental design, maximizing efficiency leads to what are called A-optimal designs (Atkinson and Donev, 1992). The essential aspect of efficiency is that it captures collinearity of each predictor with the others; if two predictors are nearly collinear their parameters cannot be estimated precisely. The variance of the predictor thus contains information about both the energy (EMP) of the predictor and its correlation with other predictors or combinations of predictors.

If a deconvolution approach is used, as it is in Dale's (1999) initial simulations, the efficiency reflects the power in estimating the shape of the hemodynamic response (Dale, 1999; Liu et al., 2001). If a basis set of hemodynamic response functions (HRFs) is used (Dale, 1999) or a single assumed HRF (Liu et al., 2001) is used, the efficiency statistic reflects the detection power of the design or the ability to correctly estimate the height parameter of each basis function.

Liu et al. (2001) demonstrated that there is an inherent trade-off between efficiency of estimating an unknown HRF shape and detection power of a signal using an assumed HRF. A single event train cannot maximize both, but pseudorandom designs whose predictors contain both high and low spectral frequencies (i.e., both blocks and isolated events) may provide a reasonable ability to estimate both the shape and the magnitude of the hemodynamic response. Birn et al. (2002) have explored this result with simulations, concluding that designs that vary rapidly between conditions (i.e., "rapid ER designs") are best for HRF estimation, whereas block designs are best for signal detection.

However, it may often be desirable to maximize HRF shape estimation power while retaining power for contrast detection. In the extreme, contrast detection with no ability to estimate the HRF (e.g., as in a blocked design) leaves ambiguity in whether results are related to a specific psychological or physiological event of interest or whether they are related to other processes that occur proximally in time (Donaldson et al., 2001). For example, comparing reading of difficult words to easier words in a blocked design could lead to differences in overall level and quality of attention during difficult blocks, as well as in specific processes that occur each time a difficult word is encountered. Designs with no HRF shape estimation power confound these types of effects. Another case where HRF shape estimation is particularly useful is in estimating differences in HRF delay onset between events (Miezin et al., 2000) and in cases where the hypothesis depends on the shape of activation, as in studies that examine maintenance of information in working memory over short delay intervals (Cohen et al., 1997).

Friston et al. (1999) gave a formulation for the standard error of a contrast across parameter estimates for different predictors, allowing for generalization of efficiency to contrasts among psychological conditions. It was subsequently generalized to account for temporal filtering and autocorrelation (see Eq. (10)). This formula easily translates into a measure of the estimation efficiency of a set of contrasts (Eq. (11)), which is an important advancement because the design parameters optimal for estimating a single event type are very different from those optimal for estimating a difference between two event types or another contrast across different conditions (Dale, 1999; Friston et al., 1999; Josephs and Henson, 1999). Considering that most researchers in cognitive neuroscience are interested in multiple contrasts across conditions within a single study, or at least the difference between an activation state and a baseline state, optimization strategies that take advantage of the ability to estimate efficiency for a set of contrasts are very useful.

*Genetic algorithms*

Genetic algorithms are a subset of a larger class of optimization algorithms, called evolutionary algorithms, which apply evolutionary principles in the search through high-dimensional problem spaces. Genetic algorithms in particular code designs or candidate solutions to a problem as a digital "chromosome"—a vector of numbers in which each number represents a dimension of the search space and the value of the number represents the value of that parameter (Holland, 1992). Analogous with genes on a biological chromosome, the elements of a design vector in a genetic algorithm are parameters specifying how to construct an object. For example, in aircraft design, three parameters might be material, curvature of the wing foil, and wing length. Values in the design vector can be continuous, as with length, or nominal, as with material. In fMRI design, the chromosome is a vector of $N$ time bins, and the values

in each slot represent the type of stimulus presented, if any, at that time. This method allows for temporal jitter at the temporal resolution of the time bins, with variable numbers of rest bins following each event of interest creating stochastic temporal variation (e.g., Burock et al., 1998). However, the resolution specified is arbitrary.

A second common feature of genetic algorithms is that selection of designs operates on numerical measures of fitness. Design vectors must be translatable into single numbers that represent the suitability of the design. In aircraft design, an engineer may want to minimize material costs and fuel consumption and maximize carrying capacity and maximum safe flying speed. These quantities must be assessed through simulation and collapsed to a single number that represents the overall suitability of a particular design specification. In fMRI, researchers may want to simultaneously maximize the efficiency of estimation of several contrasts and the efficiency of estimating the hemodynamic response for statistical reasons and minimize stimulus predictability for psychological reasons.

Three processes with analogs in evolutionary theory drive the operation of genetic algorithms: selection, crossover, and point mutation. Genetic algorithms generally start with a population of randomly generated design vectors, test the fitness of those vectors, select the best ones, and recombine the parameter values (i.e., exchange some elements) of the best designs. The designs can then be tested again, and the process is iterated until some criterion is reached. This process is analogous to natural selection: Designs are randomly rather than explicitly constructed, but the best designs from each generation are carried over to the next. On average, combinations of numbers that code for desirable features tend to spread throughout the population, and the population evolves through successive generations toward higher fitness scores.

The process of exchanging elements among successful designs also has a biological analog, which is referred to as "crossover." Although there are many ways to exchange design elements, one of the simplest is for successful (i.e., high-fitness) vectors to be paired together, as biological chromosomes are, and exchange the half of their parameters that lies above a randomly selected point along the vector. The crossover process is illustrated Fig. 1A.

A third process important in genetic algorithms is random reassignment of the values of a small number of design elements to other values, as shown in Fig. 1B. This process is analogous to point mutation in biology, which refers to the random mutation of certain components of DNA. The functional value of this process in the algorithm is to ensure that the population of designs as a whole remains heterogeneous. Crossover relies on heterogeneity in the population; essentially, the heterogeneity determines how large of a search space the algorithm covers. While it is unlikely that random mutations alone will generate high-fitness designs, random mutation in combination with crossover can be very effective.

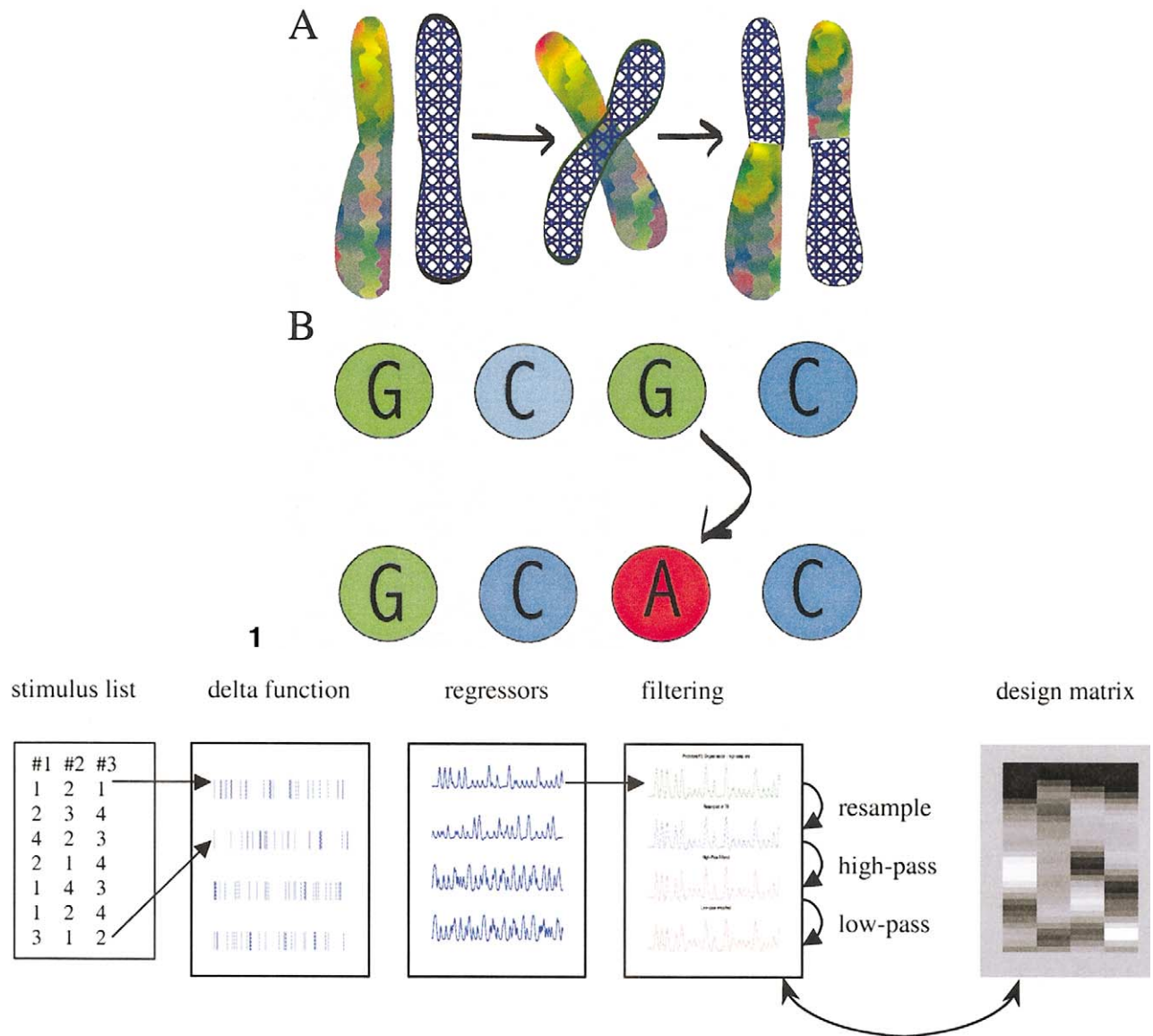Genetic algorithms have been applied over the past 40

Fig. 1. Illustration of crossover (A) and point mutation (B). In crossover, chromosomes in biology or vectors in a genetic algorithm pair up and exchange pieces at a randomly selected crossover point. In point mutation, certain genes or elements of a vector are randomly given a new value from the available range of possible values.

Fig. 2. Diagram showing the construction of the design matrix used to test efficiency from design vectors. Design vectors contain all the information necessary, with design parameters such as high-pass filter length, to reconstruct the design. The design vectors form the "genetic code" of the GA, and the best of these are intermixed between generations to create novel pseudorandom design vectors that may be yet more optimal than their "parent" vectors.
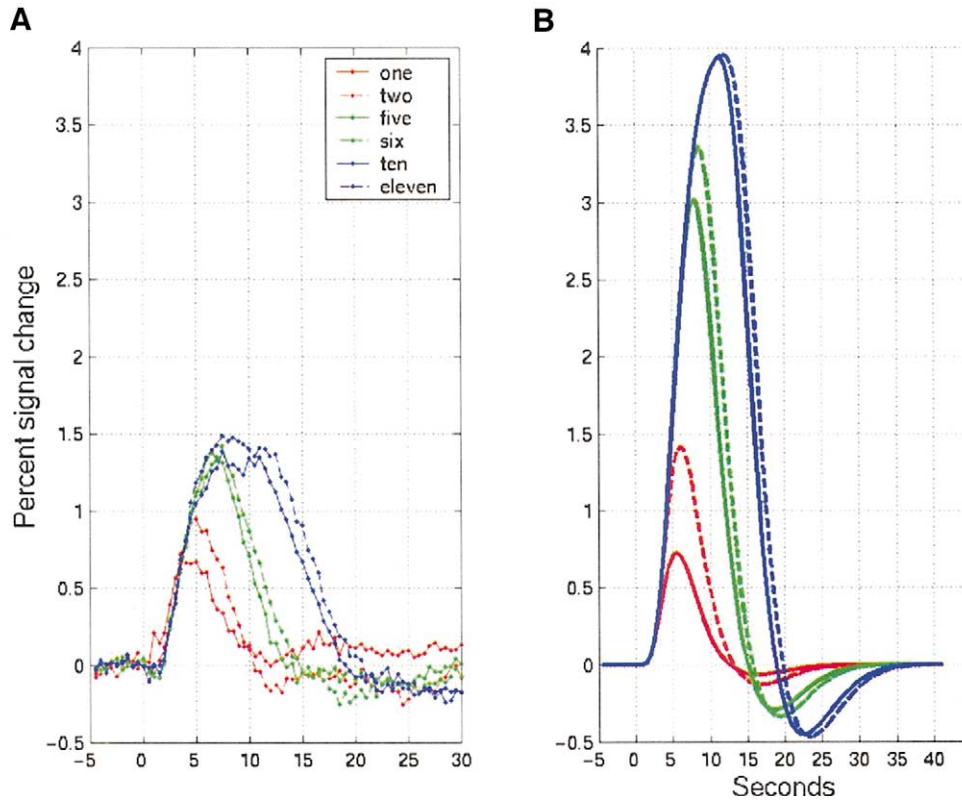
years to a number of areas of science and engineering, including structural design and statistics, and recently in biology for protein folding and synthesis of proteins and drugs (Douguet et al., 2000; Sheridan et al., 2000; Yadgari et al., 1998). They are particularly suited to problems that are difficult to solve analytically or by hill-climbing techniques such as gradient descent methods used in many nonlinear fitting algorithms and expectation maximization techniques (Dempster et al., 1977). These problems typically have a large and complex "fitness landscape"—a mapping of fitness across incrementally varying parameter values, where each parameter is a dimension of the landscape. A large search space and nonmonotonicity of fitness across many dimensions makes exhaustive search computationally prohibitive and causes hill-climbing algorithms to get stuck in local optima.
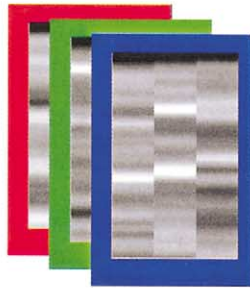
## Methods

### Linear systems approach

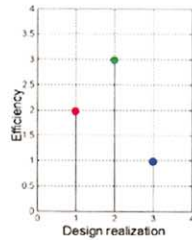Simulation is carried out in the framework of the general linear model

**A**



**B**

3
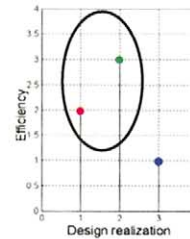
Fig. 3. Comparison of actual data for 10 subjects collected at 3 T (A) to predicted response in a linear system (B). Flickering checkerboard stimuli (120 ms) were presented, and motor responses were collected, once per second in a series of 1, 2, 5, 6, 10, or 11 stimulations, with a 30-s rest interval following the series. Data in (A) show percentages of signal change in series of each length. Linear predictions (B) were constructed by convolving a canonical HRF, normalized to the observed response height of a single checkerboard stimulation, convolved with the stimulus delta function for each series. Actual responses to long series peaked at approximately two times the height of the single stimulation response, whereas the linear convolution predicts responses over five times greater.

Fig. 4. Schematic showing the stages in GA operation. Design vectors are transformed to model matrices, which are tested for fitness. Selected designs undergo crossover, and the resulting "children" are retested iteratively until the algorithm is stopped.

$$Y = X\beta + \varepsilon, \tag{1}$$

where $Y$ is the $n$ vector of data observed at a given voxel, $X$ is the $n$-by-$k + 1$ design matrix whose columns represent the expected response, $\beta$ is a $(k + 1)$ vector of coefficients for the columns of $X$, and $\varepsilon$ is a $n$ vector of unobserved random errors; we make the usual assumption that each $\varepsilon_i$ is normally distributed with mean zero and variance $\sigma^2$. There are three typical approaches to finding $\hat{\beta}$ (an estimate of $\beta$). The first approach is to assume independent errors and use the usual least squares estimate, $\hat{\beta} = (X'X)^{-1} X'Y$. The other two approaches, prewhitening and precoloring, account for dependent errors. Let $V = \text{Var}(\varepsilon)/\sigma^2$, the correlation matrix of the errors. Both prewhitening and precoloring estimates can be written

$$\hat{\beta} = ((KX)'KX)^{-1}(KX)' KY, \tag{2}$$

where $K$ is a $n$-by-$n$ (de)correlating matrix. In prewhitening, $K$ is chosen such that $KVK'$ is the identity matrix; in precoloring $K$ is chosen such that $KVK' \approx KK'$. The prewhitening approach yields the most efficient estimator of $\beta$ with colored noise, but leads to biased estimates of $\sigma^2$ if the estimate of the noise autocorrelation is inaccurate (Friston et al., 2000a).

The noise autocorrelation estimates used in our simulations were derived empirically from a visual–motor task in which subjects were asked to press the thumbs of both hands on seeing a change in a visual stimulus (Aguirre et al., 1998), which occurred every 20 s, with images acquired every 2 s over 20 trials for a total of 200 images per subject. As in previous approaches to noise estimation (Aguirre et al., 1997; Friston et al., 2000; Zarahn et al., 1997), we assumed a stationary noise model. The data were regressed on the estimated hemodynamic response and its temporal derivative, to provide a reasonably flexible model that can account for some variation in response delay, and the residuals were used to calculate the average cross-correlation across brain voxels for each subject, giving a 50-point estimation of the autocorrelation (Box et al., 1994), which were then averaged across seven subjects to give an autocorrelation estimate specific for our imaging environment. Although we judged this to be an adequate autocorrelation model, use of the GA framework does not depend in any way on this particular estimation method; any method judged sufficiently accurate is acceptable.

*Accounting for nonlinearity*

The general linear model assumes that the response to each event is independent of and sums linearly with responses to other events. Several studies have demonstrated that nonlinear interactions in fMRI do occur (Birn et al., 2001; Friston et al., 2000; Janz et al., 2001; Vazquez and Noll, 1998), with the principal nonlinear effect being a saturation of fMRI signal when events are repeated less than 2 s apart. That is, the hemodynamic response to a second

event is less than to the first, and the estimated response height decreases still more for subsequent occurrences of the event. Boynton et al. (1996), although they report linear BOLD responses with respect to stimulus duration, found saturation effects in response magnitude. Other labs have found nonlinear saturation with stimulus presentation rate (Berns et al., 1999; Binder et al., 1994), consistent with the idea that closely spaced events produce signal saturation. An alternative way of conceptualizing the effect we term "saturation" is to say that the fMRI response to initial, brief events is *greater* than the value expected under linearity. The difference between these two conceptualizations amounts to a difference in the reference function (an impulse response for the saturation account and a block of stimuli for the alternative account); the same underlying phenomenon is described by both accounts. This effect is relatively small at long ISIs, but can have a substantial impact on designs with ISIs less than 4 s. Simulation with ISIs of less than 1 s may be completely untenable without a model of signal saturation. In previous simulations, design efficiency increases without bound as the ISI becomes very short (Dale, 1999; Friston et al., 1999; Josephs and Henson, 1999). In effect, discrete linear convolution, which is typically used to construct predicted responses, becomes meaningless at short ISIs. With continuous stimulation discretized at ISI s, the predicted response magnitude approaches infinity as ISI approaches zero, whereas the actual response magnitude is invariant with respect to the sampling resolution of the indicator function (i.e., the stimulus onset times).

In a block design with task and control periods of equal length and number of stimuli, this effect is not a problem. In continuous-stimulation block design models, where the ISI is assumed to be zero, scaling of the predictors depends on the actual sampling resolution at which the design is constructed. This resolution is often at least equal to the repetition time (TR) of volume acquisitions. Sampling at different TRs results in an equivalent scale change in the response estimates and standard errors, leaving the resulting test statistic invariant to sampling rate. However, when block design predictors (or repeated, closely spaced events) are mixed with individual events (for an example, see Liu et al., 2001, Fig. 4), the situation becomes problematic: the scaling of the block portion of the predictor relative to the individual events is arbitrary, depending on the TR, and linear convolution can yield very poor predictions of the signal response.

To give some measure of design efficiency at short ISIs, some measure of signal saturation is necessary. If the amount of saturation at a given ISI is known, the model matrix $X$ can be adjusted accordingly, and simulation can proceed. Our approach here is empirically based. In a recent study (T.D. Wager et al., manuscript in preparation), we compared trains of 1, 2, 5, 6, 10, and 11 visual stimuli (reversing-contrast checkerboards) and motor responses spaced 1 s apart. We observed that the peak magnitude for

trains of 5, 6, 10, and 11 stimuli were approximately equal, indicating that it is difficult to increase activity above this level, as shown in Fig. 3. The peak value was about twice the magnitude of the response to a single brief impulse.

In our simulations, we modified the linear convolution by adding a "squashing function" that replaced values greater than 2× the impulse response with the 2× value. Thus, the predicted response sums linearly up to twice the impulse response height, at which it is completely saturated. While nonlinear effects are clearly more complicated than the squashing model suggests and more accurate models of nonlinearity are needed, we adopt this as an interim solution because it is easily computed and captures the basic idea of signal saturation. Again, the GA framework can accommodate any model of nonlinear BOLD response.

### Measures of design efficiency

The principal measure of uncertainty in estimation is the variance–covariance matrix of $\hat{\beta}$. When the data are independent, this is

$$\text{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}. \tag{3}$$

Thus, the uncertainty in the estimates depends on the properties of the design matrix and the noise variance $\sigma^2$. When the data are dependent and (de)correlated with matrix $K$, we have

$$\text{Var}(\hat{\beta}) = \sigma^2 (Z'Z)^{-1} Z' KVK' Z (Z'Z)^{-1}, \tag{4}$$

where $Z = KX$, the filtered design matrix (Friston et al., 1999, 2000a). If we are interested in $b$ contrasts contained in a $b \times k + 1$ matrix $C$, we focus on the variance of $C\hat{\beta}$.

Assuming unit variance and substituting the pseudoinverse of $Z$ for $(Z'Z)^{-1} Z'$ (Graybill, 1976, p. 28), we have

$$\text{Var}(C\hat{\beta}) = CZ^- KVK'(Z^-)'C'. \tag{5}$$

Design efficiency is the reciprocal of an arbitrary combination of the diagonals of $\text{Var}(C\hat{\beta})$. Use of the trace corresponds to equally weighting each contrast of interest. We use a weighting vector $w$ of length $b$ whose weights represent the importance of each contrast for the overall aim of the study. The overall measure of design efficiency is

$$\xi = 1/\text{trace}\{\text{diag}(w)\text{Var}(C\hat{\beta})\}, \tag{6}$$

where $\text{diag}(w)$ is a diagonal matrix comprised of the elements of $w$. Several special cases are worth considering. If $X$ consists of a deconvolution matrix (Glover, 1999; Hinrichs et al., 2000; Menon et al., 1998), and $C$ is the identity

matrix, Eq. (7) gives the estimation efficiency for the shape of the HRF (Liu et al., 2001). If $C$ is instead replaced by a basis set of canonical HRF vectors, then the expression evaluates the efficiency of estimating the basis set (Dale, 1999). In our simulations, we use $\xi$ to measure efficiency of both a set of contrasts and the HRF deconvolution model for all trial types.

### Parameterizing a model

The genetic algorithm works by interchanging pieces of design vectors that represent the stimulation sequence during scanning. The fitness metric, however, is derived from the filtered design matrix $Z$ associated with the stimulation sequence. It is important, therefore, to parameterize the design matrix as accurately as possible, including data filtering and analysis choices (see Fig. 2).

In our parameterization of design matrices, the design vector contains one element per ISI, and the value of each element is an integer representing the stimulation condition at that time during the experiment, with 0 for rest and 1 . . . $k$ arbitrarily assigned to conditions. For example, the integer 1 in the 5th element of the design vector might represent a visual stimulation presented at the onset of the 5th acquisition, and 2 in the 10th element might represent auditory stimulation during the 10th acquisition.

Each design vector is transformed into a high-resolution design matrix whose columns are the indicator functions for the event onsets of each trial type sampled at a high time resolution (we used 100 ms) and convolved with a canonical hemodynamic response function (SPM99, http://www.fil.ion.ucl.ac.uk/spm). The design matrix is resampled at the TR, yielding the model matrix $X$. If a whitening approach is used, the model matrix is decorrelated as per $K$. If a coloring approach is used the model matrix may then be correlated via high- and/or low-pass filtering, yielding the filtered model matrix $Z$. Filtering is optional and depend on the projected analysis strategy for the study; in general, the simulation is most accurate when the design matrix is treated exactly like the design matrix of the actual analysis.

### Psychological restriction of search space

The search through possible designs may be constrained according to a number of psychological factors, including restrictions on repetition of stimuli and transitional probabilities, or counterbalancing, among trial types. In the GA framework, psychological considerations are decided upon by the experimenter and implemented as either hard or soft

Fig. 5. Efficiency plots (linear model) over ISI and temporal jitter, created by inserting a number of rest intervals at random locations into the design, for Simulation 1. The length of the simulated scan was kept constant at 480 s, and the efficiency of a difference between two event types (a [1-1] contrast) was estimated for each combination of ISI and proportion of rest intervals. (A) Results assuming linear additivity of the HRF response. Efficiency approaches infinity at short ISIs because the signal response can sum without bound. (B) Line plots of efficiency as a function of ISI under the linear model for selected proportions of rest intervals, 1, 41, and 81%.
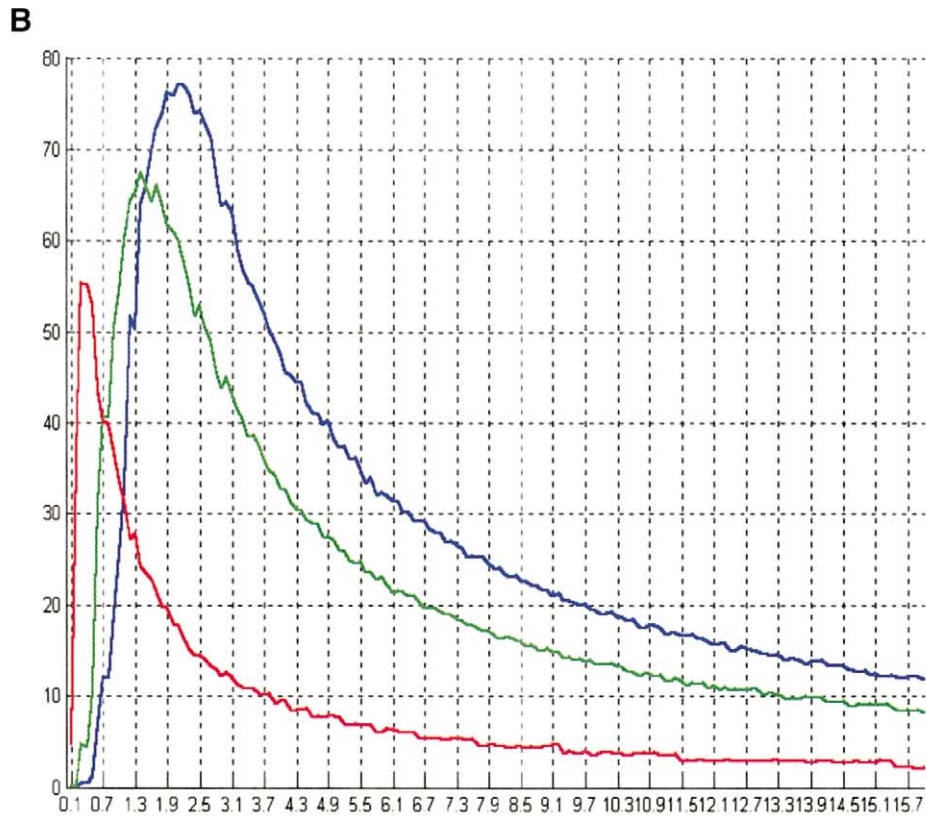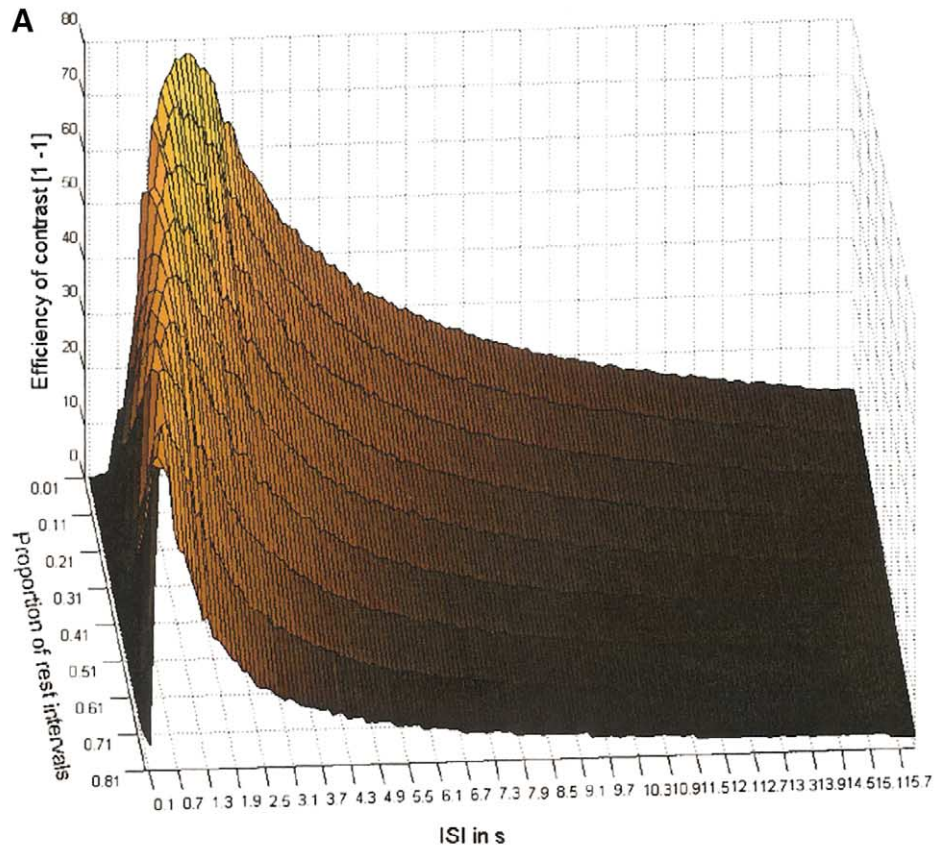
**A**



**B**



ISI in s
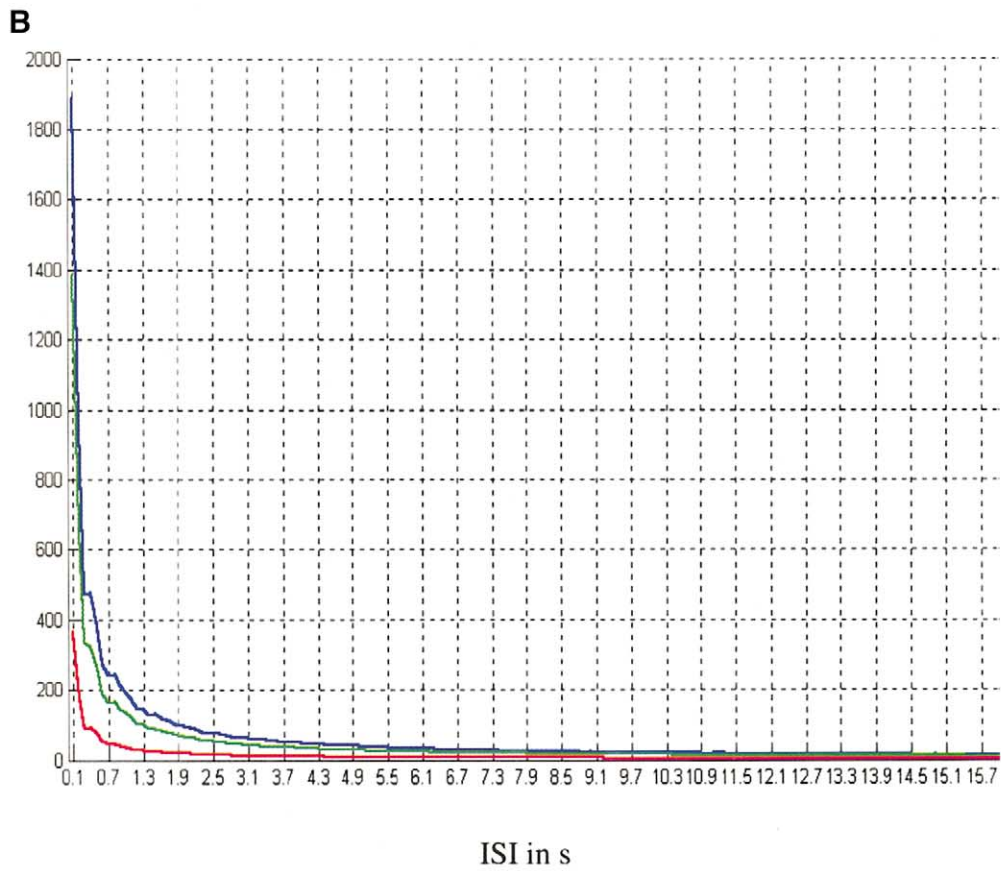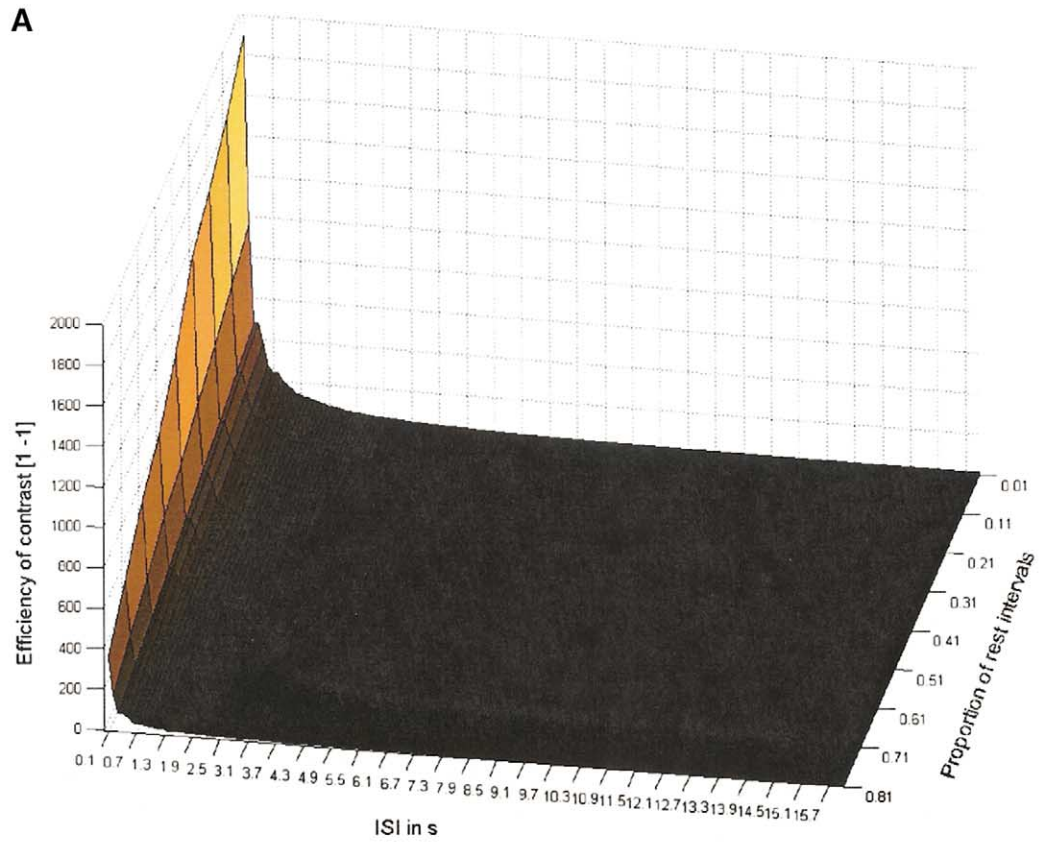
Fig. 6. Efficiency plots (nonlinear model) for Simulation 1, assuming a saturation threshold of 2× the impulse response. (A) Efficiency map over ISI and temporal jitter. The map suggests that the optimal ISI for difference detection is ~2 s, or 4 s on average between repetitions of the same event type. A greater degree of saturation than modeled would shift the peak of the curve to the right, toward longer ISIs. While the nonlinearity model is simplistic and may overestimate the amount of signal nonlinearity, it provides a starting point for obtaining reasonable estimates at all ISIs. (B) Line plots of selected levels of rest intervals for the saturation model, 1, 41, and 81%. Higher efficiencies are produced with additional randomly placed rest intervals only at ISIs less than 1.5 s.

constraints. Hard constraints are implemented by excluding design vectors that do not meet specified criteria from further consideration. Soft constraints are implemented by taking a weighted sum across scores on all fitness measures of interest to yield a single overall fitness score, which is then used as the criterion for selection.

Acceptable limits on stimulus repetition may vary depending on the psychological nature of the task. In our GA simulations, we excluded design vectors that contained more than six of the same trial type in a row anywhere in the series.

Counterbalancing of stimuli assures that the event sequence is not predictable to participants. In a perfectly counterbalanced design, the probability of each event type is independent of the event history. For counterbalancing to the first order, we consider the joint probability of trial type $j$ following trial type $i$, which is

$$p_{ij} = p_i p_j, \tag{7}$$

where $i$ and $j$ index trial type, from 1 to $k$, and $p_i$ and $p_j$ are the base rates for each trial type. If all events occur at equal rates, then $p_{ij} = 1/k^2$, and in a counterbalanced design each event type follows each other one equally often. The expected number of trials of type $j$ following type $i$ is

$$E(N_{ij}) = ap_{ij}, \tag{8}$$

where $a$ is the number of stimulus presentations. In matrix notation,

$$E(N) = app', \tag{9}$$

where $p = \{p_1 \ldots p_k\}$ is a column vector of base rates. The formula can be generalized to counterbalancing up to the $r$th order. In an $r$th order counterbalanced design, we consider the joint probability of trial type $j$ on the current trial and trial type $i$ at lag $r$, $E(N_{ijr})$. Thus the overall expected values given an $r$th order counterbalanced design can be represented by a $k \times k \times r$ three-way array $E(N)$. In practice, the expected number of stimulus presentations per condition at each time lag $r$ is the same, so $E(N_{..r})$ contains the same values for all $r$.

Given the expected number of trials for each combination of stimulus presentation orders, it is simple to compute the deviation of any design vector from perfect counterbalancing. The $k \times k \times r$ array $N$, which represents the number of observed occurrences of event pair $i$, $j$ with time lag $r$, is compared to the expected array $E(N)$ by taking the average sum of squared differences between the two arrays. This yields an overall counterbalancing score $f$ with a lower bound of 0 indicating perfect counterbalancing. The associated fitness measure $\zeta = 1 - f$ (so that higher values signify greater fitness) can be used to compare design vectors with one another and can be computed using

$$\zeta = 1 - \frac{1}{IJR} \sum_{ijr} (N_{ijr} - E(N_{ijr}))^2 \tag{10}$$

While designs with large $\zeta$ have good overall counterbalancing, we are also concerned with the worst-case counterbalancing for any given trial type. Hence we may impose a hard constraint by eliminating any designs for which max $\{(N_{ijr} - E(N_{ijr}))^2\}$ exceeds 10% of the expected number of stimuli $E(N_{ijr})$.

We also need to constrain the degree to which a design vector contains the desired frequencies of each trial type. This quantification is necessary because the recombination of design vectors in the GA does not strictly maintain the input frequencies for all design vectors. We define

$$\omega = 1 - \max|g - ab|, \tag{11}$$

where $g$ is a $k$-length vector containing the number of observed events in each condition.

### Search methods: exhaustive search and genetic algorithm

As the optimization of trial structure is a search through parameter space, the search algorithm used should depend on the size and complexity of the search space and the fitness landscape or the mapping of design fitness onto search parameters. Small search spaces can be mapped using exhaustive search, in which every combination of parameters (or a set covering the space with reasonable thoroughness) is tested, and the ultimate design is selected by using the design with the highest estimated fitness. As an example, we chose to estimate the optimal ISI and temporal jitter (introduced by inserting a random number of rest intervals into an event-related design; Burock et al., 1998) by generating random designs at a range of simulations and testing their fitness using the detection power measure described above. The results of these tests are reported in Simulation 1.

We used the GA to choose the optimal ordering given a selected ISI and number of rest intervals. The starting point for the GA was a population of design vectors created with the desired number of events in each condition in random order. Design matrices were constructed for each design vector, and fitness measures to be limited by hard constraint (including $\zeta$, $\omega$, and maximum number of repeated events, depending on the simulation) were estimated for each.

Each fitness measure was standardized by the population mean and standard deviation; this makes the metrics comparable and gives more weight to measures with small variability. A linear combination of these standardized measures that defines the overall fitness measure

$$F = w_\xi \tilde{\xi} + w_\zeta \tilde{\zeta} + w_\omega \tilde{\omega}, \tag{12}$$

where $\tilde{\xi}$, $\tilde{\zeta}$, and $\tilde{\omega}$ are the standardized efficiency, counterbalancing, and condition frequency measures, respectively, and $w_\zeta$, $w_\zeta$, and $w_\omega$ are their weights.

The fitness measure is transformed with a sigmoid (cho-

sen arbitrarily to implement a soft cutoff in fitness scores) and combined with a uniform(0,1) variate:

$$F_t = \text{rand}[0,1] + 1/(1 + e^{-\alpha F}), \tag{13}$$

where $F_t$ is the transformed fitness score, and $\alpha$ is a constant representing selection pressure or the degree of "softness" in the selection process. Design vectors with $F_t$ above the median were selected for crossover. "Selection pressure" in this context is the probability of selecting a design with $F < \bar{F}$. The case $\alpha = 0$ produces random selection (all $F_t$ are equal), and $\alpha \to \infty$ produces a step function where all designs with above-median $F_t$ also have above-median $F$. The best value for $\alpha$ was determined to be 2.1, found by systematically varying the parameter value and observing the GA's performance; specifically, the GA produced designs of highest fitness per iteration in our simulations with this value of $\alpha$. Use of random noise with the sigmoid transform allowed for some random variation in which designs were selected, preserving population variability. This principle of promoting diversity is important for systems that rely on natural or stochastic variation to produce optimal designs, and we found it to be helpful here.

Selected design vectors are paired at random and crossed over at a single randomly selected point until the population limit is reached. The portion of the vectors that lies above the crossover point is exchanged, so that the new design vectors consist of pieces of two "parent" vectors. After crossover, as in biology, a small percentage (we chose 0.1%) of the values across all lists are subject to random mutation, in which two vector elements are randomly interchanged. This process can provide additional leverage for selection, but its main function is to maintain variability in the population and prevent the system from getting "stuck" in a suboptimal solution.

Another concern is the population becoming too homogeneous. So another step taken is to introduce randomness if all lists are too correlated. Specifically, if the average correlation between lists rises above 0.6, 10% of the event codes of each design are replaced with random event codes. The event codes to be replaced are randomly selected for each design, and the best design is left intact.

To preserve the integrity of the most optimal design thus far, a separate copy is made of this design and inserted into the population without crossover or mutation. The population of new design vectors is then reevaluated for fitness, and the process of evaluation, selection, and crossover continues iteratively until a specified number of generations is reached, as shown in Fig. 4. An alternate strategy to using a fixed number of generations is to monitor convergence by plotting $F_t$ (e.g., see Fig. 9) and exiting the algorithm when the system does not improve in a specified number of generations.

## Simulations

### Simulation 1

This simulation demonstrates exhaustive search over ISIs and jitter. This simulation was designed to maximize detection efficiency for a [1 −1] contrast across two equally probable conditions and generate an efficiency surface exploring the ISI by trial density space. No filtering was used, and V determined from local data was used for this and all other simulations, except where otherwise noted. The efficiency of 100 random event sequences at each combination of ISI and jitter was estimated, and their mean efficiency was used as the population reference. ISI ranged from 0.1 to 16 s, in increments of 0.1, and jitter ranged from 1% of the total stimulation periods being rest intervals to 81% rest intervals. The TR was 2 s, with 240 acquisitions; these parameters were used for all simulations. Separate simulations were run with and without nonlinear saturation.

### Simulation 2

Previous researchers have suggested that block designs are more efficient at detecting differences between conditions than event-related ones (Dale, 1999). This simulation examines this assertion by using the GA to optimize the efficiency for a [1 −1] contrast across two equally probable conditions with no intrinsic autocorrelation model (i.e., independent noise), no filtering, and no psychological constraints. In this situation, the GA should converge on a block design. We explored this prediction further by rerunning the simulation with a highly autocorrelated noise model. The noise model was created with the autocorrelation function $1/(t/14 + 1)$, where $t$ is time in seconds, chosen to resemble typical highly autocorrelated data from our scanner (see Fig. 8C). We reasoned that increased scanner noise at low frequencies should push the optimal design into higher frequencies so that it overlaps less with the noise. For both of these simulations, TR and ISI were fixed at 2 s and no nonlinear saturation was used. Generation size was 500.

### Simulation 3

This simulation was designed to test the ability of the GA to optimize multiple efficiency measures simultaneously and to compare GA optimization versus random search through design space. The GA was set to optimize detection efficiency for a single predictor, with half of the events being the condition of interest and half being rest. The TR was 2 s and the ISI was 2 s. An identical simulation was performed using random search instead of the GA. Rather than subjecting design vectors to crossover, the random search algorithm simply rerandomized the event order in the population of design vectors and tested their fitness, saving the best vector at each iteration. The algorithm returned the best single design vector tested. Both simulations were run for equal time, each examining approximately 13,000 designs (45 generations for the GA). Generation size was 300.
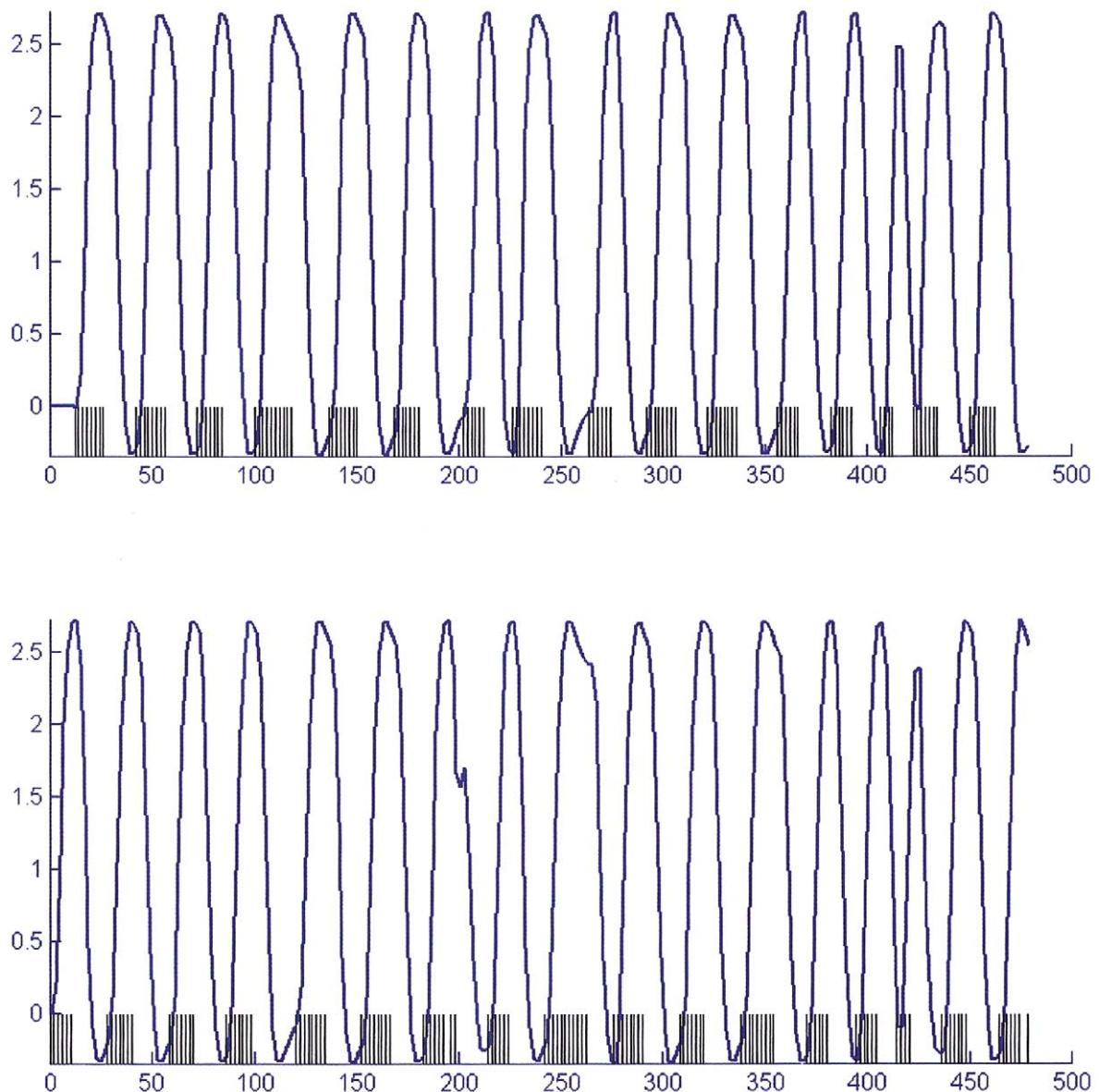
Fig. 7. Optimized predictors for the difference detection. The top panel shows condition A, and the bottom, condition B. Black lines indicate stimulus onsets.

In a second set of simulations, we used random and GA search algorithms, with the same parameters, to optimize both contrast and HRF estimation efficiency simultaneously. If there is latitude to partially circumvent the inherent trade-off between contrast estimation and HRF estimation, the GA may produce designs that are superior to randomized designs on both efficiency measures.

### Simulation 4

The purpose of this simulation was to test the ability of the GA to simultaneously optimize multiple constraints across a range of ISIs for more than two conditions. We simulated rapid event-related designs with four conditions at an ISI of 1 s and compared the performance of the

GA after 1128 generations of 300 event lists each (10 h of optimization) with groups of 1000 random designs at ISIs ranging from 0.4 to 8 s. The GA optimized (1) efficiency for a [1 1-1-1] contrast across the trial types, (2) HRF estimation efficiency across the trial types, (3) counterbalancing of all trial types up to third order, (4) 1 and 2, or (5) 1–3. Rest periods were included and maintaining the input frequencies of each condition was not specified as a priority ($w_\omega = 0$), so the GA could select the number of rest periods in the optimal design; the average time between trials of each type could vary, with a lower limit of $\sim$ 4 s. This constraint was implicit in construction of the simulation with 1 s ISI and four trial types of interest.
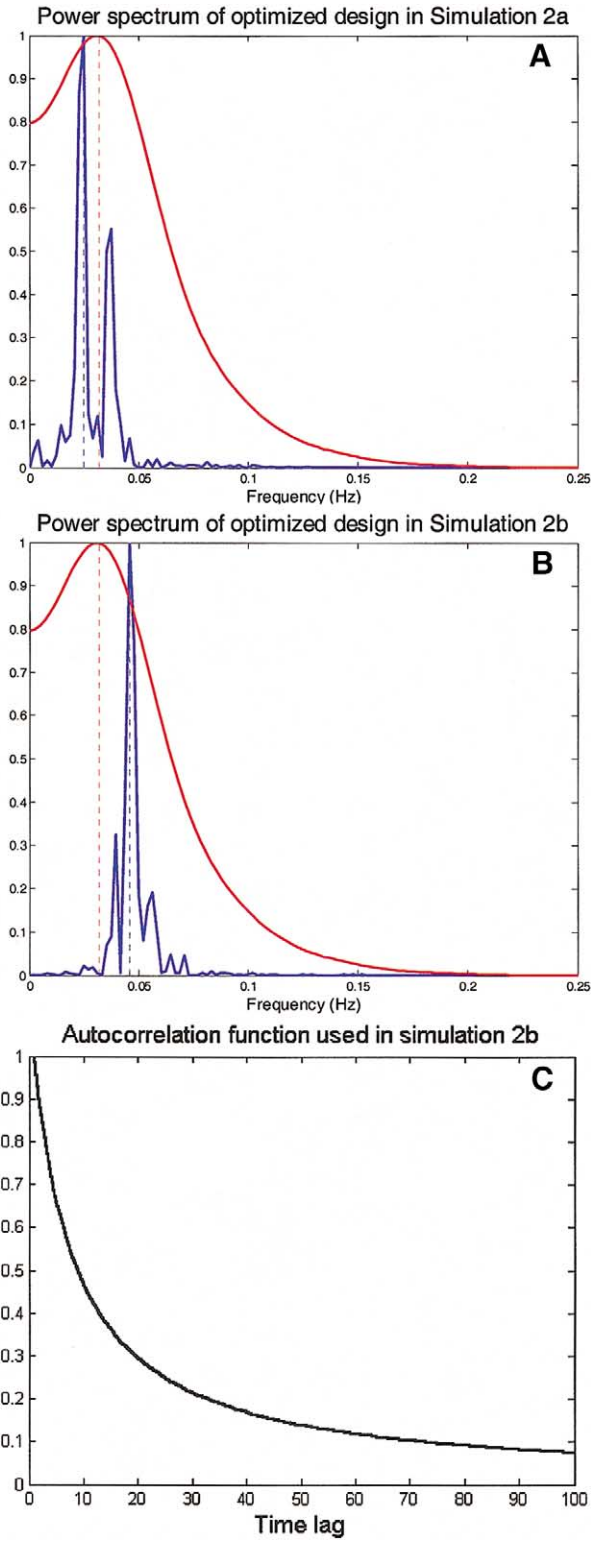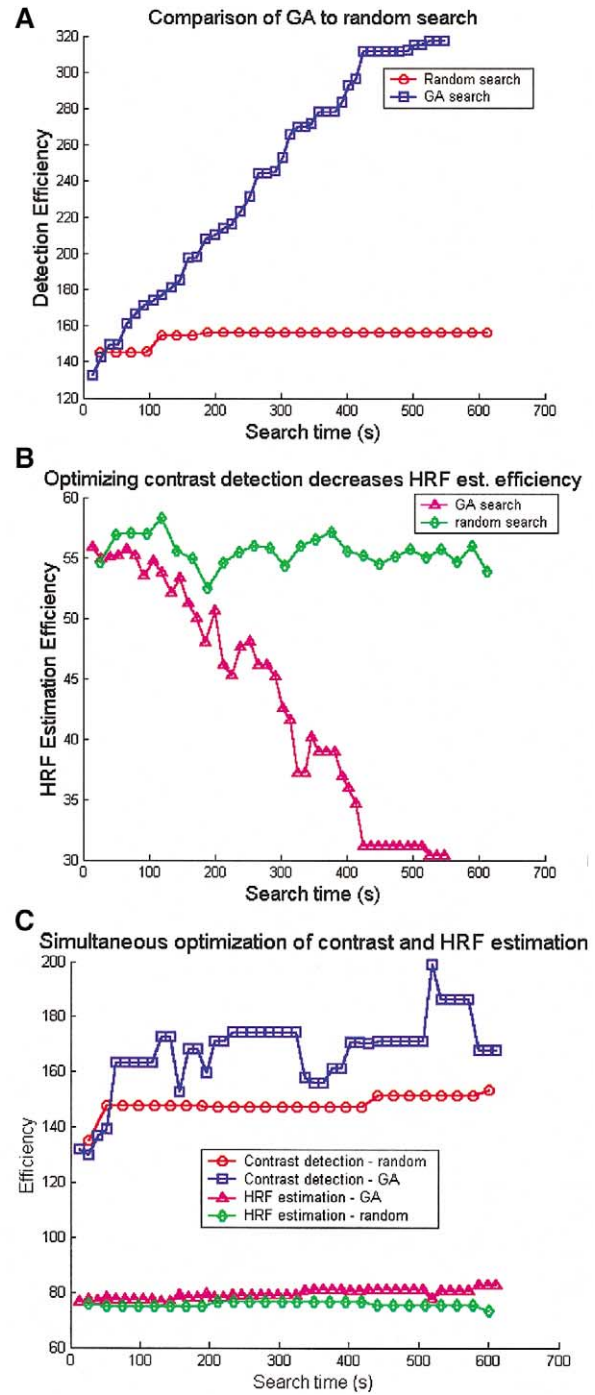
**Fig. 8.** (A) Spectral density plots of the optimized model (blue line) and the canonical SPM99 HRF (red line). (B) The same spectral profiles for a optimized model assuming high noise autocorrelation, with the autocorrelation function shown in (C). The frequency of block alternations in the evolved model increases under high noise autocorrelation, moving the power spectrum of the signal out of the range of the noise.

**Fig. 10.** Comparison of genetic algorithm (GA) with optimization by random search over a 10-min search period. (A) When viewing performance as measured by detection efficiency, the greater slope of the GA line indicates the superior performance of GA search. Both search sequences considered the same number of lists, about 13,000. Although inferior to the GA's performance, random search presents a large advantage over using a randomly generated design. The range of efficiency for 1000 random designs with the same parameter settings was 63.05–142.37. (B) Plots of hemodynamic response (HRF) detection efficiency for the same optimization period. Optimizing detection efficiency decreased HRF estimation efficiency in proportion to the gains seen in (A). HRF estimation efficiency of 1000 random designs had a range of 44.54–58.50. (C) It is possible to simultaneously optimize both detection and HRF estimation efficiency, at a cost to both. Simultaneous optimization using random search and GA search over a 10-min (~45-cycle) optimization period are shown. The GA performs better on both efficiency measures. The lines do not progress smoothly because efficiency on one measure can be sacrificed for efficiency on the other.
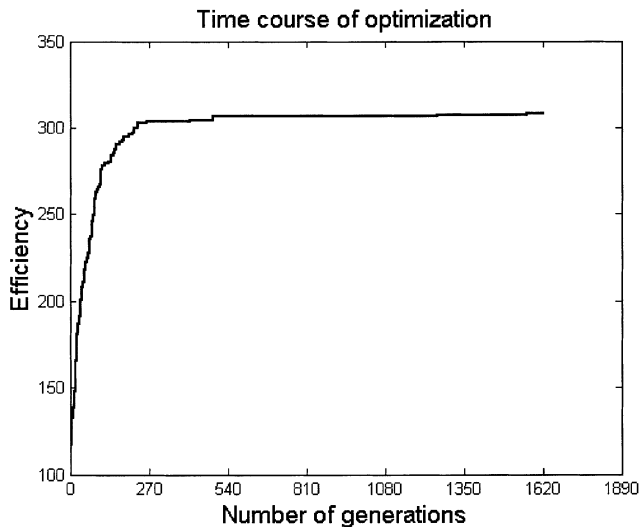
Fig. 9. The figure shows the time course of optimization for Simulation 2b. The GA converges on a block design in approximately 1 h or 2500 cycles with 300 models per cycle.

## Computation times

Several metrics might be used to measure gains in efficiency using GA or other search algorithms. One might choose to track efficiency as a function of number of cycles (generations), number of models tested, or computation time. We present simulations here in terms of computation time, annotated with description of the number of cycles. Although computation time depends on the speed of the computer system, for a given system, run time is the most direct way to compare different search parameters and algorithms, and how much time it takes to achieve a certain result is in some sense the "bottom line" in choosing a way to search. The number of models tested and the number of cycles depend on the particulars of the simulation and are not comparable across different simulations even on the same computer.

For comparability, all simulations were run on a Pentium III 700-MHz computer with 128 Mb of RAM. In 10 min of search time with a generation size of 300 designs, the GA could complete approximately 45 cycles totaling ~13,000 designs tested. These numbers also depend on the length of the designs and the complexity of the filtering and autocorrelation options, but where we compare simulation times, these variables are held constant.

## Results and discussion

### Simulation 1

The surface maps in Figs. 5 and 6 show efficiency as a function of ISI and trial density assuming signal response linearity (Fig. 5A and B) and assuming a simple nonlinear

squashing of the signal response at $2\times$ the impulse response (Fig. 6A and B). Results assuming linearity essentially replicate the findings of Josephs and Henson (1999, Fig. 2). Efficiency increases exponentially as ISI approaches zero, with dramatic increases at ISIs of less than 0.7 s. This result occurs because the signal is allowed to sum linearly without bound, and at short ISIs the signal magnitude approaches infinity.

Simulation assuming nonlinear saturation provides more reasonable estimates, although the nonlinearity model is only a rough approximation. With no rest periods, these results suggest that an optimal ISI is around 2.2 s, giving an average time between trials of the same type of 4.4 s. In this case, the random ordering of the trials provides the necessary temporal jitter, and additional rest intervals decrease power. However, at short ISIs (2 s or less), it becomes advantageous to insert rest intervals (Fig. 6), as the signal begins to saturate at these ISIs.

With greater degrees of nonlinearity, the optimal ISI grows longer. In a parallel simulation (not shown), we found that using a saturation value of $1.6\times$ resulted in an optimal ISI of 3 s or an average of 6 s between trials of the same type. Accurate models of signal nonlinearity will allow researchers to make better estimates of optimal parameter choices.

These results are for detecting differences between two conditions. If the goal is to estimate the shape of the hemodynamic response, or detect other contrasts, a different efficiency surface would be expected. The strength of this approach is that one can generate an efficiency surface specific to the aims and design requirements of a particular study, taking into account a specific autocorrelation model, filtering options, contrasts of interest, psychological design constraints, and estimation strategy. Once a particular combination of parameters is chosen from the efficiency surface, the GA can be used to further optimize the particular trial ordering.

### Simulation 2

Results show that as expected, the GA converged on a block design (Fig. 7) and that the peak of the power spectrum closely matched the peak of the power spectrum of the HRF (Fig. 8A). The peak power of the optimized design was at 0.025 Hz or 40.0 s periodicity, slightly lower in frequency than the HRF, with a second peak just above the peak frequency of the HRF. The peak power of the HRF occurred at 0.032 Hz or 31.5 s periodicity. With a high-autocorrelation noise model (Fig. 8C), corresponding to the increased low-frequency drift, we found that the solution of the GA was a block design with a higher frequency of alternation. Fig. 8B shows that the power spectrum for the design peaks at mod: 0.046 Hz, 21.8 s periodicity, substantially higher in frequency than the HRF power peak. Based on these results, the optimal design for detecting a difference between two conditions depends on the noise autocorrelation in the scanner. In particular, higher autocorrelation
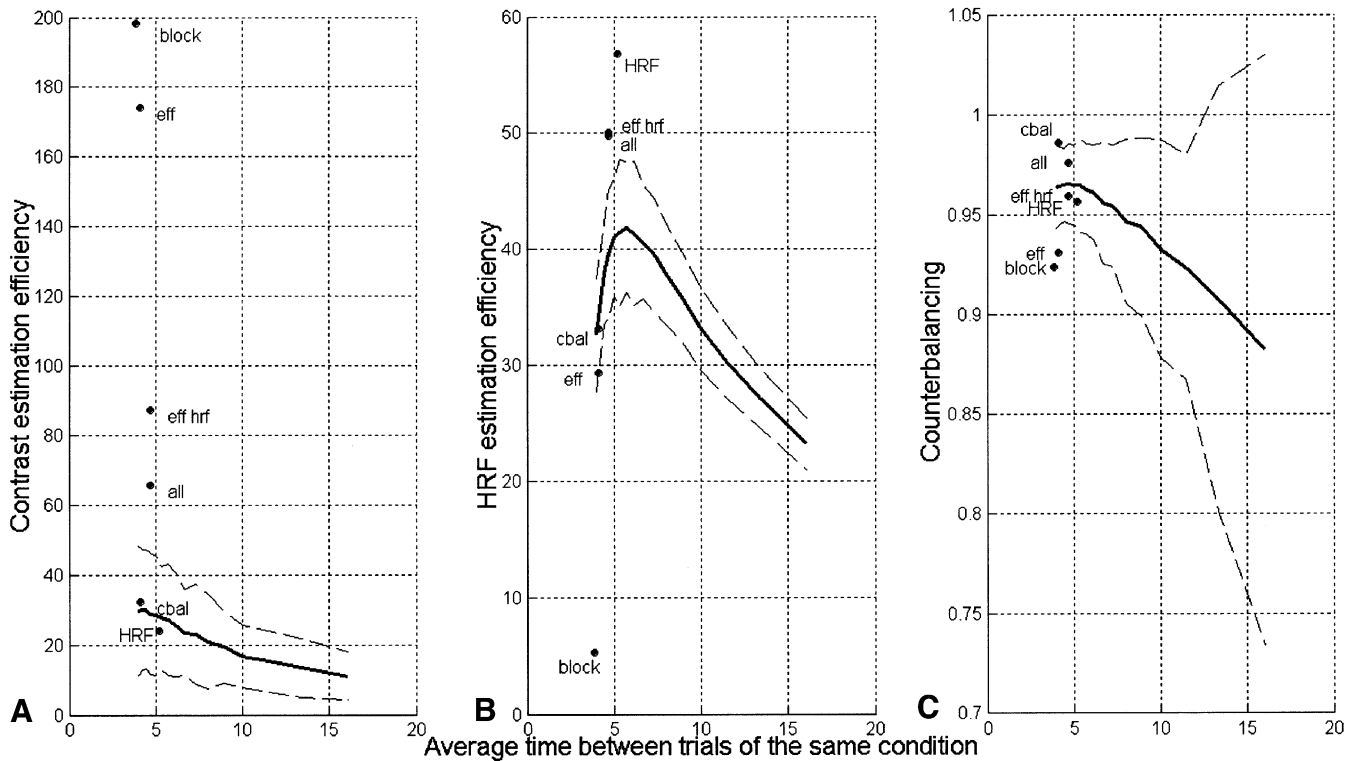
Fig. 11. Simulation 4: Optimized and block designs (black dots) compared to the mean of 1000 random sequences (black lines) on three fitness measures: (A) contrast detection efficiency (eff), (B) HRF estimation efficiency (hrf), and (C) counterbalancing (cbal). 95% confidence intervals for individual random designs at each value on the abscissa are indicated by dashed black lines. Optimizing for a particular fitness measure produces significant costs in other fitness measures. Optimizing for multiple measures produced designs that were well above random designs on all optimized measures, but not as good as is possible with selective optimization.

calls for higher-frequency designs. As Fig. 9 shows, the GA converged on the block design in approximately 250 generations.

*Simulation 3*

Fig. 10A shows detection efficiency over the search time for the GA and the random search algorithm. The GA produced much greater gains in efficiency with more iterations, indicating that it outperformed the random search algorithm. The random search algorithm was also somewhat effective; the range of efficiency for 1000 random designs was 63.05 to 142.37, with the best design substantially less efficient than produced by either GA search. However, as shown in Fig. 10B, optimizing detection efficiency comes at an inherent cost in HRF estimation efficiency, as predicted by Liu et al. (2001). In the GA simulation, HRF estimation efficiency is below the range of HRF estimation efficiency scores for random designs (efficiency ranges from 44.54 to 58.50 for 1000 random designs), indicating that the gains in contrast estimation power with the GA optimization came at the cost of decreases in HRF estimation efficiency.

An important additional question is whether the GA could optimize both detection efficiency and HRF estimation efficiency simultaneously. The results of a simulation

using both GA and random search methods are shown in Fig. 10C. With GA search, contrast efficiency improves substantially within 2 min (less than 10 generations), and the HRF estimation efficiency improves somewhat more slowly. The GA substantially outperforms random search on both measures. These results indicate that it is possible to achieve a reasonable balance between detection and HRF estimation ability. Although the optimized design is much more efficient than random event permutations on both measures, it is inferior to optimizing a single measure.

*Simulation 4*

The purpose of this simulation was to test the ability of the GA to simultaneously optimize multiple constraints (ζ, ζ, and ω̃) at a range of input ISIs. In Fig. 11, we express the results of both random simulations at a range of ISIs and the optimized designs in terms of fitness as a function of the average time between repeated events of the same type.

The results show that a short ISI—1 s, with ~ 4 s between same-type events—produces the best contrast efficiency in random designs, with a slightly longer optimal time between repetitions (~6 s) for HRF estimation efficiency. The results suggest that a compromise at ~5 s between trials of the same type will produce the best simul-

Table 1
Ratio (%) of efficiencies for selected designs compared
to random designs[a]

| Optimization type | Fitness measure | | |
|---|---|---|---|
| | Contrast eff. | HRF est. | Counterbalancing |
| Block design | 657 | 13 | 95.7 |
| Contrast efficiency | 577 | 70 | 96.5 |
| HRF estimation eff. | 80 | 136 | 99.2 |
| Counterbalancing | 107 | 79 | 102.2 |
| Contrast + HRF | 289 | 120 | 99.5 |
| Contrast + HRF + counterbalancing | 217 | 119 | 101.2 |

[a] Percentage scores are calculated as the fitness of the optimized (or block) design divided by the average fitness of 100 random designs at the optimal parameters for that fitness measure, multiplied by 100. Scores over 100 indicate an advantage over the best random design possible without search over particular random sequences. Scores less than 100 indicate significantly lower fitness than the expected fitness for the best random designs. The block design was constructed at 32 s periodicity.

taneous optimization of contrast and HRF efficiency, as indicated by the line peaks in Fig. 11A (contrast efficency) and Fig. 10B (HRF efficiency), at a small cost in counterbalancing (Fig. 11C). These results are in good agreement with similar models from other simulations (e.g., Birn et al., 2002, Fig. 2). Simultaneous optimization of contrast and HRF efficiency produced a design that was substantially better than random designs on both measures, as shown in Table 1, but was inferior to single-criterion optimization, 1:2.00 for contrast and 1:1.33 for HRF estimation. Optimization of all three fitness criteria still produced a design that was better than random designs on all three measures (Fig. 11, Table 1). This result indicates that although there are trade-offs between measures of fitness, those trade-offs are not absolute. The importance of any of the fitness criteria depends on the particular goals of the study.

*Extensions of the GA framework*

One advantage of the GA framework is that it allows for estimation of design fitness with any arbitrary design. Experimenters are not limited to modeling simple one-or two-condition designs; the choice of autocorrelation models, HRF types, and design types is arbitrary. In our lab, we have used the GA framework to develop and test mixed block/event-related designs, designs in which the stimulation is an extended epoch rather than a brief event, designs in which the conditions of interest involve dependencies among trials (e.g., task switching experiments, in which the events of interest involve changes between processing different types of stimuli, rather than the occurrence of any particular stimulus), self-paced designs, and experiments with probe and instruction periods inserted into the stimulus sequence, which change the spectral characteristics of the design. Although the possibilities are nearly limitless, the principle remains the same throughout these modifications: parame-

terize your model, choose relevant fitness measures, and let the GA find an exceptional model.

## Conclusions

Based on our simulations, it appears that it is difficult to simultaneously optimize the ability to detect a response, the ability to estimate the shape of the response, and counterbalance the ordering of events in a stimulus sequence. However, GA optimization produces designs that are much more effective by all of these criteria than random ordering of the event sequence. In addition, the GA framework produces more effective results than random search through the space of possible event orderings.

For our assumed model of the HRF and signal nonlinearity, these simulations suggest that the optimal design for detecting main effects or differences is a block design with 21–40 s periodicity, depending on the amount of noise autocorrelation, which produced approximately a 6:1 power advantage over randomized event-related designs in our simulations of difference detection across four conditions [A + B − (C + D)]. Given no other constraints, the GA framework rapidly converges to this design. For random designs, the best choice of ISI for both contrast detection and HRF estimation is one that produces trials of the same condition that are an average of 5 s apart (i.e., a 2.5-s ISI with two trial types or a 1.25-s ISI with four trial types). As nonlinear saturation of the signal response increases, the optimal ISI increases as well. Temporal jitter is recommended if the design involves only a single event type, if only main effects are of interest, or if the ISI is less than 2 s (for two-condition designs). However, use of an optimized, pseudorandom trial order can improve on random designs even at the optimal ISI.

Optimization of fMRI trial design is a difficult problem, with a number of considerations that depend jointly on the characteristics of fMRI signals, the characteristics of the brain response elicited, and the psychological goals of the study. Because of the great variation in types of design and study goals, it is difficult to come up with a simple set of rules that will produce the optimal design in all situations. The GA takes a step toward solving this problem by allowing researchers to specify designs with great flexibility, potentially capturing many of the idiosyncrasies of a study and optimizing the statistical and nonstatistical properties of the design with respect to the particular concerns of the study.

## Acknowledgments

## References

Aguirre, G.K., Zarahn, E., D'Esposito, M., 1997. Empirical analyses of BOLD fMRI statistics. II. Spatially smoothed data collected under null-hypothesis and experimental conditions. NeuroImage 5 (3), 199–212.

Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998. The variability of human, BOLD hemodynamic responses. NeuroImage 8 (4), 360–369.

Atkinson, G.L., Donev, A.N., 1992. Optimum Experimental Designs. Clarendon Press, Oxford.

Berns, G.S., Song, A.W., Mao, H., 1999. Continuous functional magnetic resonance imaging reveals dynamic nonlinearities of "dose-response" curves for finger opposition. J. Neurosci 19 (14), RC17.

Binder, J.R., Rao, S.M., Hammeke, T.A., Frost, J.A., Bandettini, P.A., Hyde, J.S., 1994. Effects of stimulus rate on signal response during functional magnetic resonance imaging of auditory cortex. Brain Res. Cogn. Brain Res 2 (1), 31–38.

Birn, R.M., Cox, R.W., Bandettini, P.A., 2002. Detection versus estimation in event-related fMRI: choosing the optimal stimulus timing. NeuroImage 15 (1), 252–264.

Birn, R.M., Saad, Z.S., Bandettini, P.A., 2001. Spatial heterogeneity of the nonlinear dynamics in the fmri bold response. NeuroImage 14 (4), 817–826.

Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J., 1996. Linear systems analysis of functional magnetic resonance imaging in human VI. J. Neurosci 16 (13), 4207–4221.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1994. Time Series Analysis: Forecasting and Control, third ed. Prentice Hall, Englewood Cliffs, NJ.

Buckner, R.L., 1998. Event-related fMRI and the hemodynamic response. Hum. Brain Mapp 6 (5–6), 373–377.

Burock, M.A., Buckner, R.L., Woldorff, M.G., Rosen, B.R., Dale, A.M., 1998. Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. NeuroReport 9 (16), 3735–3739.

Cohen, J.D., Perlstein, W.M., Brower, T.S., Nystrom, L.E., Noll, D.C., Jonides, J., Smith, E.E., 1997. Temporal dynamics of brain activation during a working memory task. Nature 386 (6225), 604–608.

Dale, A.M., 1998. Optimal experimental design for event-related fMRI. Hum. Brain Mapp 8 (2–3), 109–114.

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

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J.R. Stat. Soc. Ser. B Method. 39, 1–22.

D'Esposito, M., Zarahn, E., Aguirre, G.K., 1999. Event-related functional MRI: implications for cognitive psychology. Psychol. Bull. 125 (1), 155–164.

Donaldson, D.I., Petersen, S.E., Ollinger, J.M., Buckner, R.L., 2001. Dissociating state and item components of recognition memory using fMRI. NeuroImage 13 (1), 129–142.

Douguet, D., Thoreau, E., Grassy, G., 2000. A genetic algorithm for the automated generation of small organic molecules: drug design using an evolutionary algorithm. J. Comput. Aid. Mol. Des. 14 (5), 449–466.

Friston, K.J., Josephs, O., Rees, G., Turner, R., 1998. Nonlinear event-related responses in fMRI. Magn. Reson. Med. 39 (1), 41–52.

Friston, K.J., Josephs, O., Zarahn, E., Holmes, A.P., Rouquette, S., Poline, J., 2000a. To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. NeuroImage 12 (2), 196–208.

Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000b. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. NeuroImage 12 (4), 466–477.

Friston, K.J., Zarahn, E., Josephs, O., Henson, R.N., Dale, A.M., 1999. Stochastic designs in event-related fMRI. NeuroImage 10 (5), 607–619.

Glover, G.H., 1999. Deconvolution of Impulse response in event-related BOLD fMRI. NeuroImage 9, 416–429.

Graybill, F.A., 1976. Theory and Application of the Linear Model. Duxbury Press, Belmont, CA.

Hinrichs, H., Scholz, M., Tempelmann, C., Woldorff, M.G., Dale, A.M., Heinze, H.J., 2000. Deconvolution of event-related fMRI responses in fast-rate experimental designs: tracking amplitude variations. J. Cogn. Neurosci. 2 (Suppl 2), 76–89.

Holland, J.H., 1992. Genetic algorithms. Sci. Am. 267 (1), 66–72.

Janz, C., Heinrich, S.P., Kornmayer, J., Bach, M., Hennig, J., 2001. Coupling of neural activity and BOLD fMRI response: new insights by combination of fMRI and VEP experiments in transition from single events to continuous stimulation. Magn. Reson. Med. 46 (3), 482–486.

Josephs, O., Henson, R.N., 1999. Event-related functional magnetic resonance imaging: modelling, inference and optimization. Phil. Trans. R. Soc. Lond. B Biol. Sci. 354 (1387), 1215–1228.

Liu, T.T., Frank, L.R., Wong, E.C., Buxton, R.B., 2001. Detection power, estimation efficiency, and predictability in event-related fMRI. NeuroImage 13 (4), 759–773.

Menon, R.S., Luknowski, D.C., Gati, J.S., 1998. Mental chronometry using latency-resolved functional MRI. Proc. Natl. Acad. Sci. USA 95, 10902–10907.

Miezin, F.M., Maccotta, L., Ollinger, J.M., Petersen, S.E., Buckner, R.L., 2000. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. NeuroImage 11 (6 Pt 1), 735–759.

Rosen, B.R., Buckner, R.L., Dale, A.M., 1998. Event-related functional MRI: past, present, and future. Proc. Natl. Acad. Sci. USA 95 (3), 773–780.

Sheridan, R.P., SanFeliciano, S.G., Kearsley, S.K., 2000. Designing targeted libraries with genetic algorithms. J. Mol. Graph. Model. 18 (4–5), 320–334 525.

Vazquez, A.L., Noll, D.C., 1998. Nonlinear aspects of the BOLD response in functional MRI. NeuroImage 7 (2), 108–118.

Yadgari, J., Amir, A., Unger, R., 1998. Genetic algorithms for protein threading. Proc. Int. Conf. Intell. Syst. Mol. Biol. 6, 193–202.

Zarahn, E., Aguirre, G.K., D'Esposito, M., 1997. Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. NeuroImage 5 (3), 179–197.