ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 21 (2004) 1690–1700

# A comparison of methods for characterizing the event-related BOLD timeseries in rapid fMRI

John T. Serences*

*Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD 21218, USA*

Received 28 August 2003; revised 6 December 2003; accepted 9 December 2003

**Information about the shape and temporal duration of the blood oxygenation level dependent (BOLD) response can inform both functional neuroanatomy and psychological theory. However, the BOLD response evolves over 20 s or more, making it difficult to distinguish the unique characteristics of the response evoked by temporally adjacent stimuli. Fortunately, event-related BOLD signals can be extracted given that there is adequate variance in the distribution of inter-stimulus intervals (ISI). Unfortunately, the ISI distribution that yields the highest statistical efficiency is not always optimal from a psychological perspective; variability in the stimulus timing may complicate the interpretation of neuroimaging data in terms of underlying cognitive operations. In the present paper, Monte Carlo simulations are used to evaluate two techniques for estimating the event-related BOLD timeseries—event-related averaging and deconvolution using the Ordinary Least Squares estimate —with respect to maintaining acceptable levels of statistical power and experimental validity. While the unbiased deconvolution technique more robustly estimates the shape of the BOLD response functions, both methods succeed in accurately re-producing known differences between evoked BOLD responses when the stimulus ordering is randomized. However, the deconvolution method is more effective at preserving differences when there are sequential dependencies in the stimulus presentation order and restricted ISI distributions are used; particularly if the second of two sequentially dependent stimuli is omitted on some portion of the trials. Importantly, the successful re-production of the evoked BOLD response using restricted ISI distributions often maximizes the ability to make psychologically valid experimental conclusions.**


*Keywords:* fMRI; BOLD; Simulation; Time course; Deconvolution; Event-related average

## Introduction

With advances in analysis techniques over the last several years, researchers using event-related functional magnetic resonance imaging (fMRI) are now able to evaluate theoretical questions based on the time course, and not just the spatial distribution, of neural activity as revealed indirectly by changes in the BOLD signal (Buckner et al., 1998; Burock et al., 1998; Corbetta et al., 2000; Dale and Buckner, 1997; Miezin et al., 2000; Yantis et al., 2002). As a result of this increased focus on event-related timeseries, several methods have been developed to allow changes in the BOLD signal evoked by different events to be separated. The fundamental computational problem is the 'sluggishness' of the BOLD signal, which generally reaches a maximum over the first 4–10 s and then gradually diminishes over the next 10–20+ s (Bandettini, 1993; Boynton et al., 1996; Buckner et al., 1996; Fransson et al., 1999). Consequently, if two events are presented less than approximately 20 s apart, the BOLD response to the first event will temporally overlap the BOLD response to the second event, making it difficult to properly attribute changes in the BOLD signal to one event or another.

Early event-related fMRI studies dealt with this problem by using inter-stimulus intervals (ISIs) approaching 20 s, thus minimizing the interference between the BOLD responses to temporally adjacent events (Botvinick et al., 1999; Buckner et al., 1996). However, this approach is not ideal for at least two reasons. First, the large temporal separation between events means that only a limited number of trials may be presented in a given unit of time. The limited number of trials gives rise to a perpetual statistical problem in fMRI—because of the low signal-to-noise ratio, detecting differences between the BOLD responses to different events is often quite difficult. Second, using long ISIs to temporally isolate the BOLD response to adjacent events is not ideal from a psychological standpoint. For example, suppose a subject was performing a delayed match-to-sample task in which they are to remember a sample stimulus for some delay interval and then determine if the test stimulus matches the sample. If the sample and test stimuli were separated by 20+ s to allow estimation of the BOLD response to the sample stimulus, a subject's memory for the sample may be significantly degraded, making the neuroimaging data difficult to interpret in terms of the underlying cognitive function.

In addition, well-controlled cognitive psychology (and neuroimaging) experiments depend on successfully isolating specific mental operations of interest. Consequently, long ISIs are often undesirable because the mental operations being performed by a subject during any given task become increasingly unclear as the

* Department of Psychological and Brain Sciences, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218. Fax: +1-410-516-4478.
*E-mail address:* serences@jhu.edu.
**Available online on ScienceDirect (www.sciencedirect.com.)**

1053-8119/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2003.12.021

amount of time in between adjacent stimuli is increased (see Stark and Squire, 2001). Thus, event-related fMRI paradigms involve compromise between several competing factors. On one hand, temporally sluggish BOLD responses motivate using ISI distributions containing some events that are separated by long temporal interval (Buracas and Boynton, 2002; Dale, 1999); on the other hand, statistical and psychological considerations dictate that different events should be presented as close together in time as possible (Liu et al., 2001; Wager and Nichols, 2003).

Fortunately, the intrinsic properties of the BOLD response allow both of these constraints to be at least partially satisfied. Critically, BOLD responses to temporally adjacent events can be adequately described as a linear time invariant system as long as the ISI does not fall below several seconds (Boynton et al., 1996). That is, the evoked response to a given event type will result in an approximately equivalent contribution to the overall BOLD signal regardless of the current signal level when the event is presented (Boynton et al., 1996; Dale and Buckner, 1997). While non-linearities develop as events are presented closer and closer together in time, the assumption of linearity has been shown to be adequate for intervals as short as 2 s under the proper circumstances (Burock et al., 1998; Dale, 1999; Dale and Buckner, 1997). Because event-related BOLD responses can generally be modeled as a linear system, the problem of determining the unique response elicited by a particular event type involves solving a series of linear equations (Fig. 1). The critical constraint on solving any system of linear equations is that there must be as many unique equations as there are unknowns. As depicted in Figs. 1a, b, if the ISI is fixed, there are more unknowns than unique equations, thus, it is impossible to reconstruct the basis functions that make up the BOLD timeseries. However, if the ISIs are randomly jittered, then a unique equation can be derived for each estimation point along the BOLD response function (Figs. 1c, d, see section on deconvolution techniques below for more details; see Burock et al., 1998).

To make matters even more complex, experimental paradigms sometimes require sequentially dependent events; such is the case with the classic delayed match-to-sample task or Posner's spatial cueing task, designs that are frequently used in cognitive neuroscience (Corbetta et al., 2000; Courtney et al., 1997; Hopfinger et al., 2000; Sala et al., 2003). In both of the paradigms, event type A (e.g., the 'sample stimulus') is always followed by event type B (e.g., the 'test stimulus'). In such experiments with sequential dependencies, as opposed to experiments with randomized event ordering, estimating the BOLD response to a particular event is problematic because different events might have different effects on the time course in a given brain region. For instance, if event A

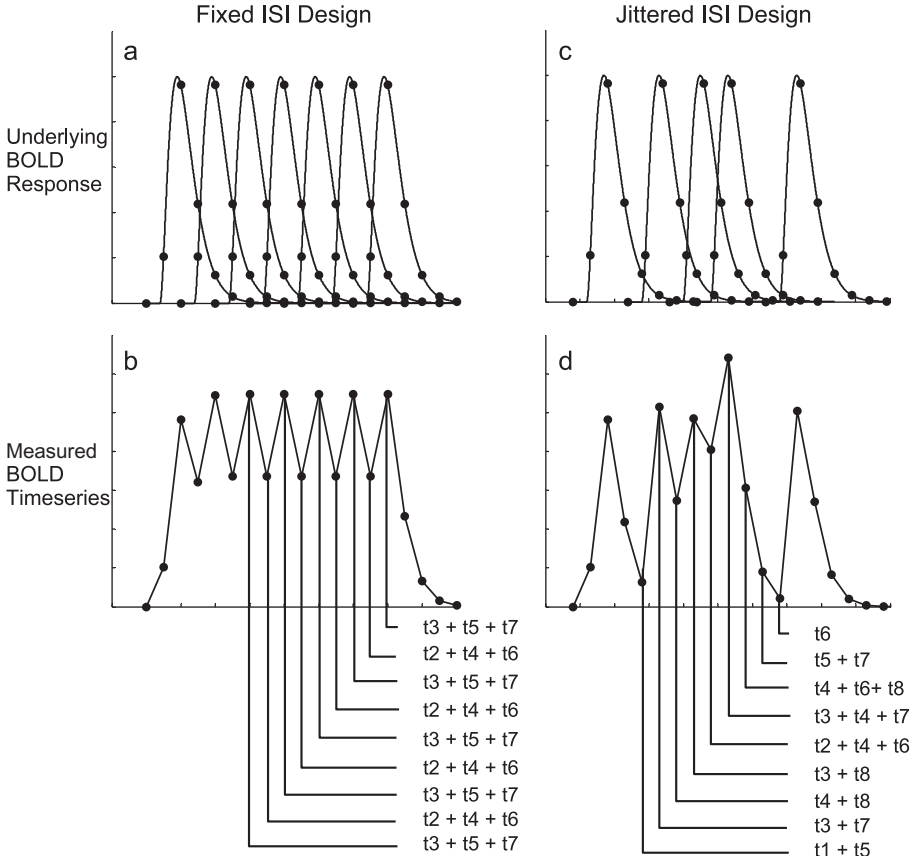


Fig. 1. Schematic of the effects of using fixed and random ISIs on the measured BOLD response. The underlying BOLD response function is composed of 8 timepoints for simplicity (t1–t8). (a) Using a fixed ISI design will produce BOLD responses that overlap consistently in time, resulting in a lack of variance in the measured BOLD time series when it is discretely sampled during a scanning session (b). From a linear systems point of view, a fixed ISI will result in fewer unique equations than unknowns, making it impossible to uniquely reconstruct the underlying BOLD response function. If the ISI is jittered (c), then the measured BOLD response will be relatively complex, resulting in the ability to form more unique linear equations than there are unknowns (d). Adapted and reprinted with permission from (Miezin et al., 2000).

consistently evokes a small BOLD response in visual cortex, and event B consistently evokes a large response in the same cortical area, problems may arise when event A is always followed by event B. Specifically, to the extent that the procedure employed to estimate the evoked BOLD response is imperfect, the estimated response to A might be artificially inflated due to the heightened response to event B (Ollinger et al., 2001a, 2001b; Woldorff, 1993).

In the present paper, simulated experiments are carried out using various event ordering and temporal constraints. Comparisons are then made between two different techniques that are commonly used to determine the event-related BOLD time course; event-related averaging and deconvolution of the BOLD timeseries using the Ordinary Least Squares estimate (OLS). Both of these methods involve selective averaging of the BOLD response to a given stimulus type, but only the deconvolution technique takes into account the potential contamination of the event-related BOLD signal by the presentation of temporally adjacent stimuli. Accordingly, the simulation results show that when BOLD responses to adjacent stimuli do not systematically interact (e.g., randomized stimulus ordering), then reasonable statistical conclusions can be drawn based on either event-related averaging or deconvolution techniques. However, as previously noted in the event-related potential literature, sequential dependencies in the stimulus presentation order may lead to severe distortions in the event-related averages (e.g., Woldorff, 1993). When restricted ISI distributions are used—which are often ideal from psychological standpoint—the distortions in the event-related averages become increasingly severe. Thus, the present results demonstrate that event-related averaging should only be used in rapid fMRI paradigms when a sufficient degree of randomization can be introduced into the stimulus presentation order. If this condition is not met, then deconvolution techniques should be used to maximize the ability to accurately characterize the event-related BOLD response while simultaneously satisfying psychological constraints on the experimental design.

## Materials and methods

### *Event-related averaging*

Computing event-related averages of the BOLD timeseries is a method adopted from the event-related potential literature and is carried out by extracting the timeseries for a specified number of timepoints before and after the onset of each event type. The extracted timeseries for each event type are then averaged together—as in all averaging techniques, it is assumed that any random noise will cancel out over repeated trials and the remaining waveform will be dominated by the signal evoked by that particular event type. In the present simulations, the baseline (or 0% signal change level) for the event-related averages was defined as the average of the signal level during the 6 s before event onset (e.g., Kirino et al., 2000; Yantis et al., 2002). This baseline choice is somewhat arbitrary, but the results presented here are robust across various baseline intervals.

### *Modeling the BOLD timeseries*

In most event-related fMRI paradigms, the BOLD signal may be adequately described as a linear time-invariant system (Boynton et al., 1996; Friston et al., 1995a, 1995b) and thus modeled at a given time ($t$) using the General Linear Model (GLM) of the form:

$$\mathrm{s}(t) = b_0 + x_1(t)b_1 + x_2(t)b_2 \cdots + x_{n-1}(t)b_{n-1} + \varepsilon(t) \quad (1)$$

where $n$ is the number of variables to be estimated, $x$ is a user-defined stimulus response function for a given variable, and $b$ is a scalar beta weight that minimizes the error between $\mathrm{x}_n(t)$ and $\mathrm{s}(t)$. Written in the more compact matrix form this equation is:

$$\boldsymbol{s} = \mathbf{X}\boldsymbol{b} + \varepsilon \quad (2)$$

where $\boldsymbol{s}$ is the observed BOLD timeseries vector, $\mathbf{X}$ is the *design matrix* that describes the relationship between stimulus presentations and the observed data, and $b$ is a vector of model parameters that describes the magnitude of the relationship between $\mathbf{X}$ and $\boldsymbol{s}$. The design matrix $\mathbf{X}$ has a row for each timepoint in the BOLD signal ($s$) and a column for each estimated variable. The vector $\boldsymbol{b}$ contains a scalar value for each corresponding column in $\mathbf{X}$. The least squares solution that minimizes the error between $\mathbf{X}$ and $\boldsymbol{s}$ is found by inverting the design matrix in the form:

$$\boldsymbol{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T s \quad (3)$$

The proto-typical approach to estimating the timeseries in fMRI is to convolve a fixed canonical response function (e.g., a gamma or Gaussian function) with a boxcar model of the stimulus sequence (Cohen, 1997). The convolved model timeseries is inserted into a column in $\mathbf{X}$ that represents the expected response to a particular stimulus across the entire timeseries. Each column in $\mathbf{X}$ is used to represent a different variable and is associated with a scalar value in the corresponding row of vector $\boldsymbol{b}$ that scales the canonical input function to minimize the error between $\mathbf{X}$ and $\boldsymbol{s}$.

This approach has proven to be extremely useful in evaluating the extent to which the timeseries in a given voxel corresponds to an expected response function. However, the use of a canonical response function has several drawbacks. First, the most commonly used response function was derived based on the BOLD response in V1 to a flickering checkerboard (Boynton et al., 1996). However, deviations from this pattern have been observed in various brain regions, as well as within different subjects in the same brain regions (Aguirre et al., 1998). Therefore, the validity of applying a standard response function to all voxels in the brain is sometimes questionable.

A second related problem arises because a single column in $X$ represents the expected BOLD response to all presentations of a given stimulus type and the regression procedure produces a single scalar beta weight that describes the relationship between the canonical response function and the actual timeseries. Therefore, the estimated response is a static scaled version of the input response function, which contains relatively little information about the temporal characteristics of the BOLD response. Again, any deviations from this input response function will result in a poor model fit and/or an inaccurate

characterization of the temporal profile of the event-related BOLD timeseries.

One way to circumvent these problems is to use the GLM to estimate the actual BOLD response in each voxel separately, without relying on the convolution of a canonical response function and a boxcar model of the stimulus presentation sequence (see, B. Ward, http://afni.nimh.nih.gov/afni/docpdf/3dDeconvolve.pdf; Hinrichs, 2000; Glover, 1999; Dale and Buckner, 1997, see also: Burock and Dale, 2000; Dale, 1999). This deconvolution approach is implemented by altering the design matrix $X$; instead of a single column in $X$ representing the expected response to a given event over the entire timeseries, a separate column of $X$ is created for each timepoint along the expected event-related BOLD response for each event. A '1' is placed in each row of the design matrix in the column corresponding to the point along the BOLD response that is to be estimated (Fig. 2). This method only assumes that the response to a given stimulus type is consistent over repetitions and that the number of estimated timepoints adequately accounts for the duration of the evoked BOLD response. With the design matrix specified in this way, inverting the design matrix will result in an estimate $b$ for each point along the BOLD response for each variable entered into the model.

### *Multicollinearity and deconvolution*

A limitation of the GLM in general is that the design matrix $X$ must be inverted to find the estimated values of $b$. In order for the inversion to have a unique solution, each column in $X$ must be linearly independent. More formally, no column in $X$ can be represented as a linear combination of the other columns—this requirement is met if every row in the row-reduced echelon matrix equivalent to $X$ contains at least one non-zero entry. When this condition is met, the matrix is said to have full rank. Problems of linear dependence arise because of multicollinearity in the design matrix, which occurs when there is a consistent relationship between the temporal ordering of different stimulus events. For example, consider a delayed match-to-sample task where the sample stimulus is presented at time $t$ and is always followed at time $t + 1$ by a test stimulus. If the evoked BOLD response to the sample stimulus in (say) frontal cortex is assumed to last more than one timepoint, then the change in the BOLD signal due to the sample stimulus and the change due to the test stimulus would be impossible to disentangle because the two events are perfectly correlated in time.

The general problem of multicollinearity highlights the importance of introducing temporal jitter between various stimulus events (e.g., Figs. 1c, d). In practice, if a fixed ISI that is shorter than the duration of the evoked BOLD response must be used in a paradigm with sequential dependencies, the only way to accurately reconstruct the BOLD response to different events is to use partial trials (Ollinger et al., 2001a, 2001b). In a partial trial paradigm, the second event in a sequential series is sometimes omitted (on approximately 25–40% of the trials). For instance, in a delayed match-to-sample task with a fixed delay interval, omitting the test stimulus on a subset of the trials will sufficiently reduce multicollinearity to a point where the evoked response can be uniquely estimated for the sample stimulus. Thus, partial trial experimental designs must be used when temporal jitter is prohibited by experimental considerations.

### *Simulations*

#### *Simulation parameters*

Monte-Carlo simulations were performed to evaluate two different methods of extracting event-related BOLD timeseries—event-related averaging of the BOLD timeseries and deconvolution of the evoked response using the GLM. These two techniques were used to analyze the data from three different experimental designs commonly used in cognitive neuroscience. In the first type of experiment, the temporal ordering of the stimulus events was independent; each event type was equally likely to be preceded and followed by every other event type. The second and third experiments were variations on sequentially dependent designs, in which the order of the event presentations is constrained. In the second experiment, event type A was always followed by event type B (termed a compound trial experiment). In the third experiment, the second of the two sequentially dependent events was omitted on 30% of the trials (termed a partial trial experiment, Corbetta et al., 2000, 2002; Ollinger et al., 2001a, 2001b).

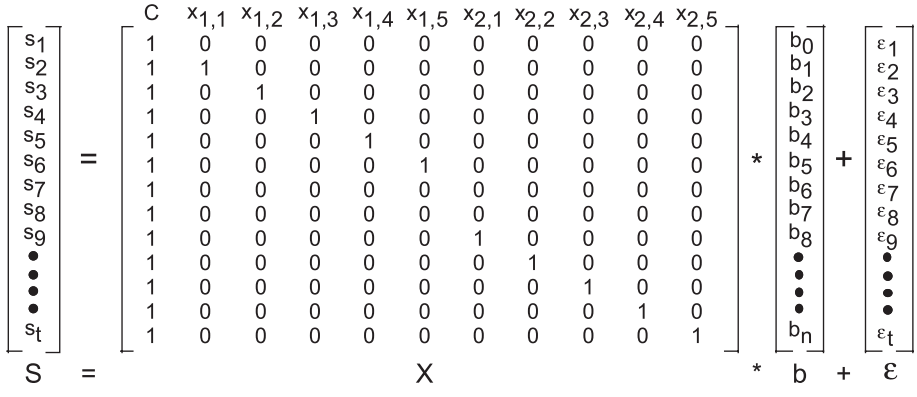


Fig. 2. Schematic of a design matrix used for estimating the evoked BOLD response to different events in the GLM framework. $S$ corresponds to the measured signal, which is modeled by multiplying each column in $X$ with a corresponding beta weight in vector $b$ such that the error ($\varepsilon$) at each timepoint is minimized. $T$ denotes the number of timepoints to be modeled and $n$ denotes the number of variables to be estimated. Each estimated timepoint is modeled with a 1 in the row of $X$ that corresponds to relative timing of the stimulus presentation. The row labeled $C$ in the design matrix represents the models constant term. In this example, the responses to events $x_1$ and $x_2$ are modeled at the first five timepoints after event onset. In practice, the number of TRs modeled following each event typically accounts for 12–24 s following the onset of each stimulus (see Ollinger et al., 2001b).

Each of the simulated experiments was evaluated using three different ISI distributions (Fig. 3). The ISI distributions are approximately exponential, with different minimum and maximum ISI values; exponential distributions have been shown to yield more efficient experimental designs than uniform ISI distributions (Dale, 1999; see also: Buracas and Boynton, 2002; Wager and Nichols, 2003). However, the present simulations produce the same qualitative data patterns across a variety of ISI distributions.

All simulations were carried out using code written in the Matlab programming language (version 6.1). A run length of 256 s was used, which included 20 s of 'rest' at the beginning and the end of each run so that the response to the last event would not be truncated. Each simulated experiment consisted of 15 iterations (or 'subjects'), and each iteration consisted of 12 runs. The signal consisted of gamma functions (delta = 2, tau = 1.25, sampling rate = 10 Hz) that were linearly summated at the onset time of each event. Zero-mean Gaussian white noise (amplitude = 14) and temporally correlated noise, consisting of 1 Hz (heart rate) and 0.2 Hz (respiration) sin waves, was added to the signal timeseries (lower frequency noise below 0.05 Hz that is commonly observed in the BOLD signal was not added because these frequencies are typically removed before data analysis). The simulated timeseries were then sampled at 1 second intervals (TR = 1 s) before event-related averaging and deconvolution analyses were carried out.

Event-related averaging was carried out by defining a temporal window extending from 6 TRs before event onset to 20 TRs after event onset and extracting the BOLD signal during this interval for every occurrence of a given event type. The extracted timeseries were then averaged together; the 0% signal change baseline was defined as the mean signal level during the 6 TRs preceding event onset. The deconvolution timeseries was created by entering a regressor into the design matrix of a GLM for the TR corresponding to the time of event onset and for each of next 19 TRs.

Four different event types (A, B, C, D) were defined and each of the events evoked a gamma response with an amplitude of (1, 3, 1, 1), respectively. For purposes of notation, the different event types will be referred to by letter with the amplitude of the response in parentheses (e.g., A(1), B(3), etc). In the simulated independent experiments, event ordering was pseudo-randomized. In the compound trial simulations, event type A(1) was always followed by event type B(3), and event type C(1) was always followed by event type D(1). In the partial trial simulations, events B(3) and D(1) were omitted on 30% of the trials (Ollinger et al., 2001a, 2001b). Because the run length was fixed, fewer instances of each event type were presented as the mean ISI increased. Thus, for ISI(1), ISI(2), and ISI(3), there were 18, 9, and 6 occurrences of each of the four trial types in a single experimental run, respectively. By design, 30% fewer occurrences of events B(3) and D(1) were present in a run of a partial trial experiment.

To evaluate the statistical power of each technique, comparisons were made between the evoked responses to each of the different event types. Specifically, the response to event A(1) was compared to response B(3), and the response to event C(1) was compared to the response to event D(1). Given the input functions in the simulated timeseries, event B(3) should evoke a larger response than event type A(1). Events C(1) and D(1) should evoke equivalent responses. The evoked responses to each event were compared at each timepoint using repeated measures *t* tests with a threshold of $P < 0.05$. Thus, the effectiveness of each method for computing an evoked BOLD response is reflected in the extent to which the relationship between the reconstructed timeseries matches the known relationships between the input response functions. It is worth noting that one would rarely analyze differences between two BOLD signals using point-wise *t* tests due to the inflated probability of false positive results; however, correcting for multiple comparisons was ignored here for simplicity because the arbitrarily chosen noise levels render the statistical analyses appropriate for comparison purposes only.

Various measures have been proposed to evaluate the statistical efficiency of evoked BOLD response measurements; however, these methods are limited to cases in which the response was estimated using the GLM (e.g., Dale, 1999). Thus, to directly compare the effectiveness of event-related averaging and deconvolution in the present study, the simulated experiments were repeated 50 times to generate an estimate of the reliability of the *t* values produced by the point-wise comparisons between different event types. It should be noted that because the noise level was fixed but arbitrary, the exact levels of significance between different events (e.g., A(1) and B(3)) is also arbitrary; however, relative comparisons between levels of significance are still valid.

## Results

### *Independent events*

Fig. 4 depicts the results from a single simulated experiment with independent event ordering for each of the three different

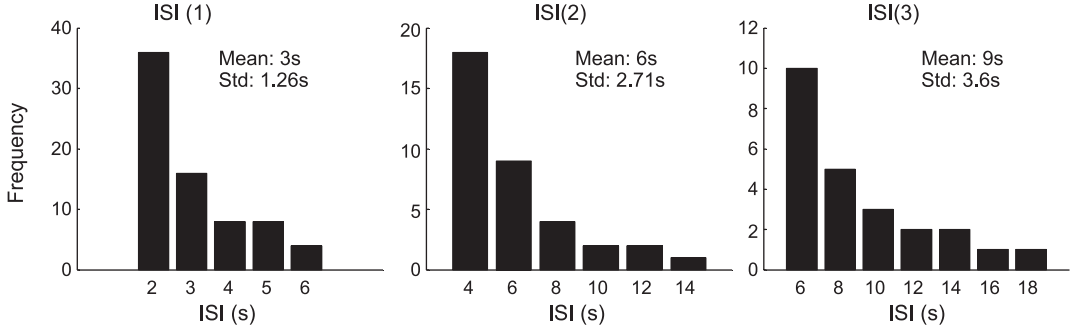


Fig. 3. Histograms representing the three ISI distributions used in the simulated experiments. ISI(1) represents a fairly restricted range of ISIs, ISI(2) represents a distribution that is more representative of a typical cognitive neuroscience experiment, and ISI(3) represents a distribution with a relatively large mean and standard deviation. See Dale (1999), Buracas and Boynton (2002) and Wager and Nichols (2003) for more information on optimizing ISI distributions.

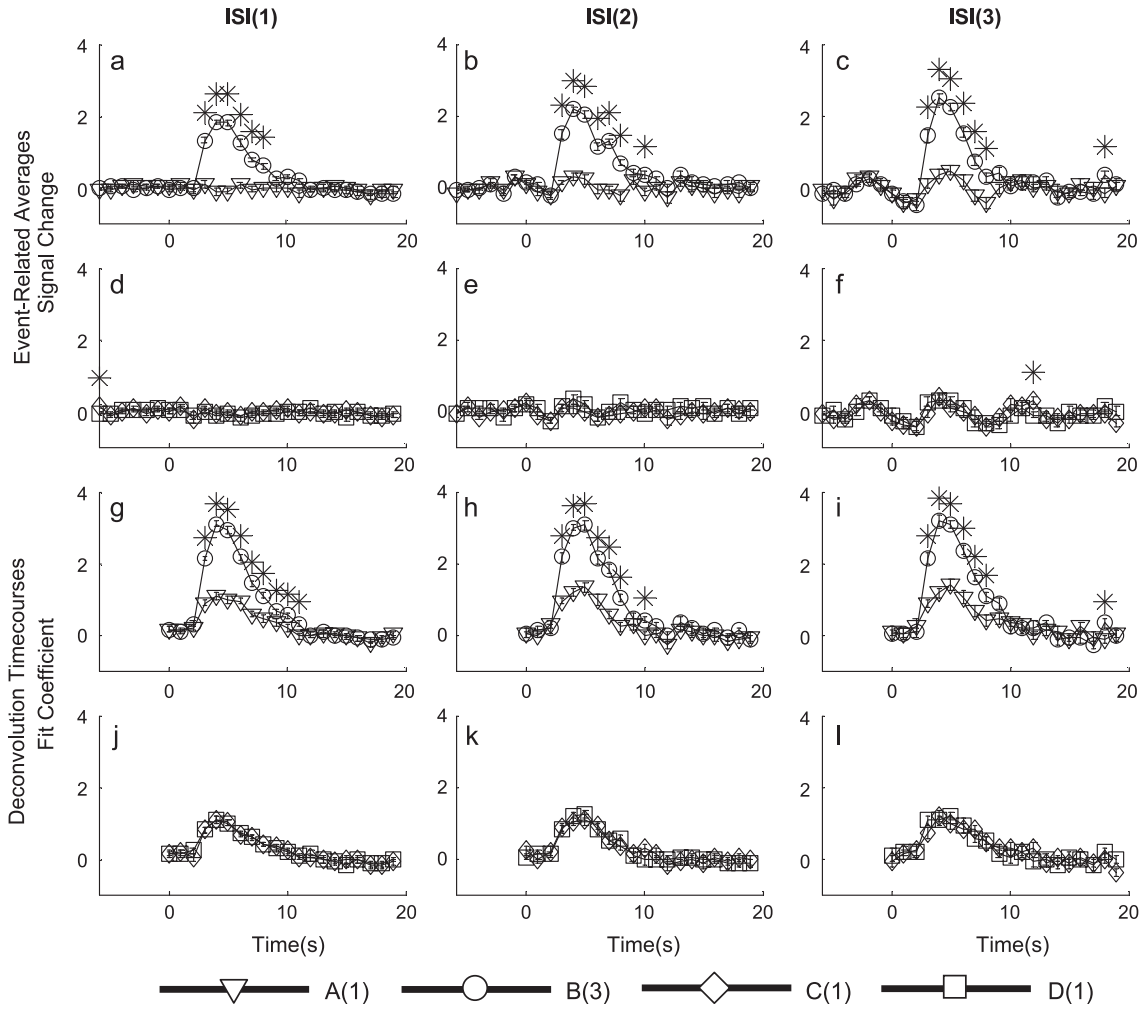


Fig. 4. Simulation results from a single experiment in which the four event types were presented in a pseudo-randomized order. Each of the three columns represents the results produced at each of the three ISI distributions. The figure legend notes the possible evoked response amplitudes that could precede and follow each event (the response amplitude to the specific event is shown in bold face type). (a–f) Event-related averages comparing the response to events A(1) with B(3) (a–c) and the response to events C(1) with D(1) (d–f). (g–l) Complementary time series plots using the deconvolution method to estimate fit coefficients (beta weights) at each of the 20 timepoints following event onset. Asterisks indicate significant differences between the timeseries as revealed by a repeated measures $t$ test at each timepoint with a threshold of $P < 0.05$. The results show that both event-related averaging and deconvolution techniques accurately reproduced the differences between A(1) and B(3), and successfully reproduced the lack of a difference between the identical events C(1) and D(1). Error bars ± 1 SEM.

ISI distributions. Event-related averaging solutions are shown in Figs. 4a–f and the deconvolution solutions are shown in Figs. 4g–l. The input response functions for events A(1), C(1), and D(1) all had an amplitude of 1, whereas the input response for event B(3) had an amplitude of 3. Thus, if the relationships between the event-related timeseries were reconstructed perfectly, the only significant difference should arise between events A(1) and B(3); events C(1) and D(1) should produce identical responses. Both the event-related averages and the deconvolution timeseries produced this expected result; there was a significant difference observed between events A(1) and B(3) across all ISIs, and no (or few) differences observed between events C(1) and D(1).

While the same statistical inferences would be drawn using either event-related averaging or deconvolution, the exact shape of the evoked timeseries was more robustly reproduced using the deconvolution technique (Woldorff, 1993). That is, the event-related averaging method led to reasonable statistical results, but the form of the response functions was somewhat distorted. For instance, a comparison of panels 4e and 4k reveals no significant differences between events C(1) and D(1) in either panel; however, the event-related averages are essentially flat. They do not reproduce the amplitude (1) response of the input HRF function (as seen in Figs. 4j–l).

*Compound trials*

Fig. 5 presents the results from a single simulated experiment when sequential dependencies are introduced into the experimental design (compound trials). In this situation, event A(1) is always followed by event B(3). Event C(1) is always followed by event D(1); however, event C(1) may be preceded by D(1) or by B(3).

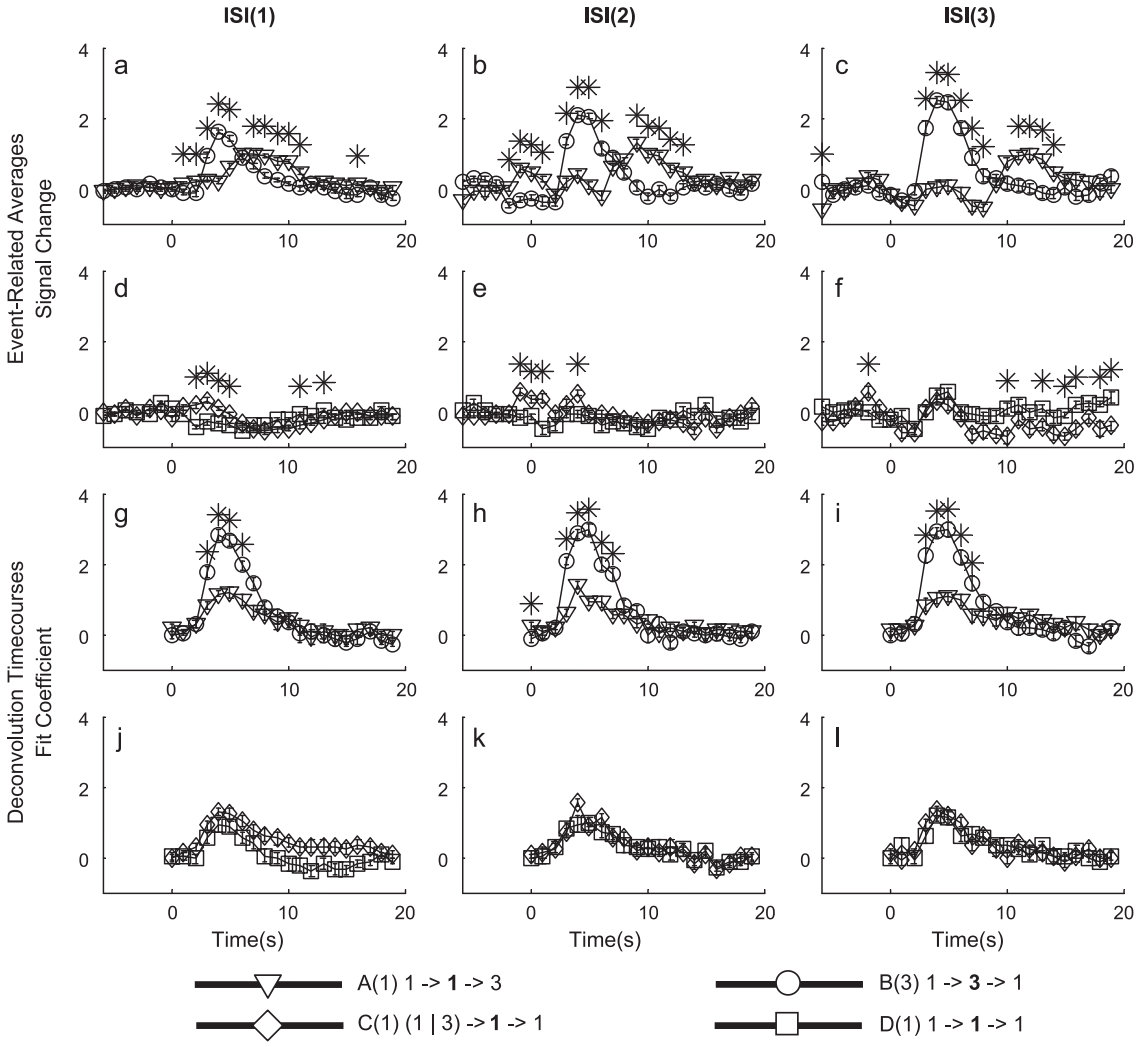


Fig. 5. Simulation results from a single experiment in which event A(1) was always followed by event B(3) and event C(1) was always followed by event D(1). The figure legend notes the possible evoked response amplitudes that could precede and follow each event (the response amplitude to the specific event is shown in bold face type). Importantly, event C(1) could be preceded by either event B(3) or event D(1). (a–f) Event-related averages comparing the response to events A(1) with B(3) (a–c) and the response to events C(1) with D(1) (d–f). (g–l) Complementary time series plots using the deconvolution method to estimate fit coefficients (beta weights) at each of the 20 timepoints following event onset. The results reveal severe distortions in the event-related average plots, particularly when the standard deviation of the ISI distribution is small (e.g., panel a). In contrast, the deconvolution technique produced accurate results across all ISI distributions; however, there is a notable increase in the standard error of the estimates at ISI(1) (panels g,j). Error bars ± 1 SEM.

Under these conditions, the event-related averages become severely distorted (Figs. 5a–f); with the overall distortion increasing as the standard deviation of the ISI distribution decreases (see also Fig. 7). For example, at ISI(1) and ISI(2), the peak event-related response to event A(1) and event B(3) are shifted significantly out of phase. At ISI(3), the peak response to event A is slightly more in phase with response B(3), but maintains an artifactual 'sustained' component because it is always followed by the high amplitude B(3) (Fig. 5c).

The heightened response to event B(3) also manifests in a difference between the responses to events C(1) and D(1), which should be identical. However, the response to event C(1), which is preceded by event B(3) on approximately 50% of the trials, is erroneously inflated (see Figs. 5d, 7d). This contamination of the event-related average to event C(1) is most evident at ISI(1), and largely dissipates as the standard deviation of the ISI distribution increases.

In contrast to the event-related average plots (Figs. 5a–f), the deconvolution timeseries estimated using the GLM method was remarkably robust (Figs. 5g–l). There are some signs of distortion at ISI(1), and the standard error of the estimates increased as the standard deviation (and mean) of the ISI distribution decreased. However, the relationship between the estimated timeseries fairly accurately reflects the relationship between the input functions across all ISI distributions.

### *Partial trials*

Depicted in Fig. 6 are the results from a single simulated experiment with sequential dependencies and partial trials. Even

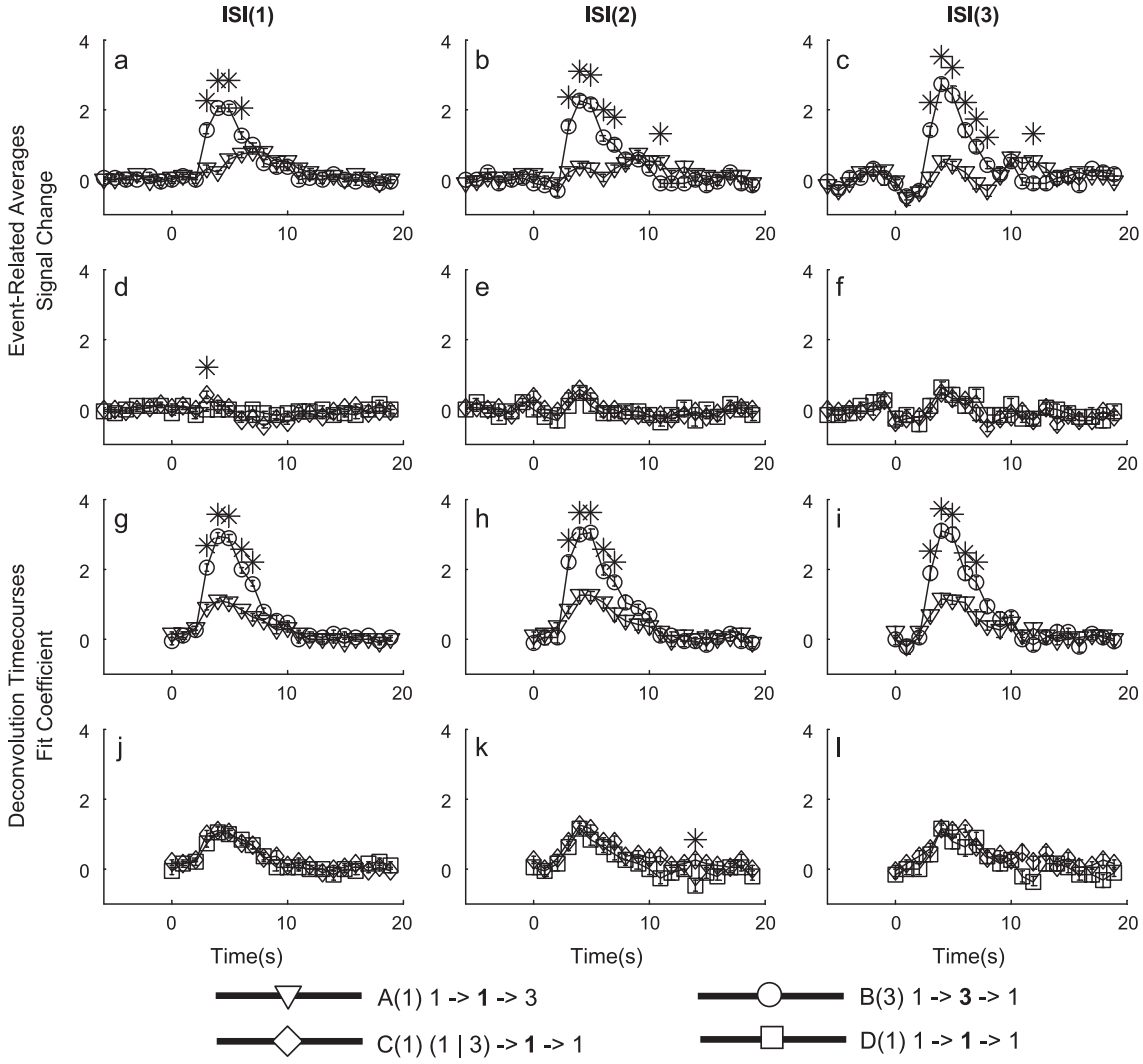


Fig. 6. Simulation results from a single experiment with partial trials; that is, event A(1) was followed by event B(3) on 70% of the trials, and event C(1) was followed by event D(1) on 70% of the trials. The figure legend notes the possible evoked response amplitudes that could precede and follow each event (the response amplitude to the specific event is shown in bold face type). (a–f) Event-related averages comparing the response to events A(1) with B(3) (a–c) and the response to events C(1) with D(1) (d–f). (g–l) Complementary time series plots using the deconvolution method to estimate fit coefficients (beta weights) at each of the 20 timepoints following event onset. The use of partial trials attenuated the distortions observed in the event-related averages that were so prominent; however, at ISI(1) and ISI(2), phase distortions between events A(1) and B(3) are still evident. The deconvolution technique produced accurate results across all ISI distributions. Error bars ± 1 SEM.

though B(3) was omitted after the presentation of A(1) on 30% of the trials, significant distortions were still evident in the event-related averages (e.g., Figs. 6a–c). As in the compound trial experiment, the deconvolution timeseries that was estimated using the GLM were robust across ISIs in the partial trial experiment (Figs. 6g–l).

*Direct comparison of event-related averaging and deconvolution*

To compare the accuracy of event-related averaging and GLM approaches with respect to re-constructing the evoked BOLD response, each of the three experiment types was repeated 50 times. The mean $t$ value across the 50 iterations was computed at each timepoint for comparisons between A(1)–B(3) and C(1)–D(1). Fig. 7 depicts the mean absolute value of the $t$ score at each timepoint for each condition. For both event-related averages and deconvolution timeseries, independently randomizing the stimulus presentation order produced the most reliable results. That is, there were significant differences where differences should have been present (e.g., between A(1) and B(3)) and no differences when the responses should have the same (e.g., between C(1) and D(1)). For restricted ISI distributions, the partial trial experiments produced more robust results compared to the compound trial experiments; however, as the standard deviation of the ISI distribution increased, the statistical power of the compound trial method approached that of the partial trial method. This convergence was especially apparent in the deconvolution timeseries; although the partial trial approach produced higher average $t$ values at ISI(1) and ISI(2), the compound method

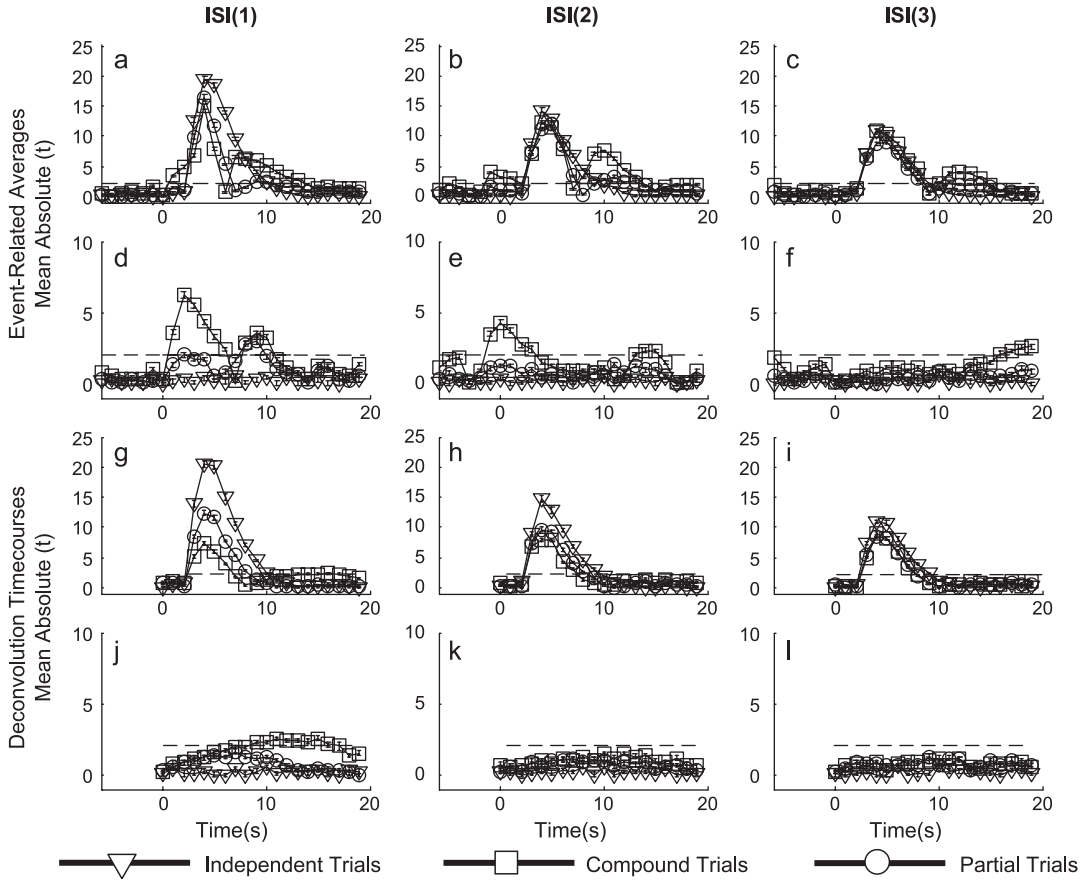


Fig. 7. To evaluate the consistency of the simulation results presented in Figs. 5, 6, and this figure, each simulated experiment was repeated 50 times. The *t* values at each timepoint comparing A(1) with B(3) and C(1) with D(1) were then averaged across the 50 iterations of each experiment. (a–f) Average *t* values for the event-related averages comparing the response to A(1) with B(3) (a–c), and comparing the response to C(1) with D(1) (d–f). (g–l) Corresponding average *t* values computed on the deconvolution time courses. Only events A(1) and B(3) should have produced significant differences, and these differences should have roughly been restricted to timepoints 5–10 post-stimulus. The extent to which other timepoints and conditions were significantly different reflects error in re-creating the input response functions. Horizontal dashed line represents *t* value corresponding to $P < 0.05$.

was just as reliable at ISI(3) (compare circles and squares in Figs. 7g and i).

## Discussion

The present simulations evaluate the effectiveness of two different techniques for constructing event-related representations of the BOLD timeseries; event-related averaging of the BOLD timeseries and deconvolution of the event-related response using the GLM. When the temporal ordering of the stimulus presentations can be adequately randomized, event-related averaging produced statistical results that are comparable to deconvolution; however, if event ordering is completely dependent, then event-related averaging can produce severely distorted timeseries (Woldorff, 1993). Therefore, the present simulations suggest that the deconvolution method is preferable across a wider variety of experimental paradigms.

### *Event-related averaging versus deconvolution with sequentially dependent events*

The deconvolution method more robustly characterized the BOLD timeseries when there were sequential dependencies in the stimulus presentation order. The deconvolution technique is able to more accurately reconstruct the timeseries because each point along the estimated timeseries represents the solution to a system of linear equations that minimizes the difference between the response functions specified in the design matrix and the actual timeseries. When there are overlapping BOLD responses to temporally adjacent stimuli, the corresponding design matrix in the deconvolution GLM will also contain overlapping regressors for each event, thereby correcting for contamination of the evoked BOLD response by temporally adjacent stimuli. In contrast, event-related averaging simply sums the BOLD signal from all specified timepoints and divides by the number of events that went into the summation. Consequently, if there are systematic differences in the evoked BOLD responses to temporally adjacent events, event-related averaging techniques will always be biased to some extent (Dale, 1999; Woldorff, 1993).

The simulations also support previous findings showing that partial trial paradigms are statistically preferable to compound trial paradigms when the standard deviation of the ISI distribution is restricted (Ollinger et al., 2001a, 2001b). Introducing partial trials effectively increases the amount of jitter between events; this increased temporal jitter results in increased estimation consistency when parceling out the variance in the BOLD signal to be

attributed to a given regressor in the design matrix. The increased jitter due to partial trials has a similar effect when computing event-related averages; there is less overlap between the BOLD responses to temporally adjacent events and thus less distortion in the resulting waveforms. However, the advantages of the partial trial approach diminished as the standard deviation of the ISI distribution increased (for example compare Figs. 7g–i). This occurred because fewer events were presented per unit time, which necessarily results in some loss of statistical power given that there is noise in the system. Remarkably though, the loss of power that accompanies presenting fewer stimuli is offset by the advantages of decreasing the temporal overlap between adjacent BOLD responses, at least for the more restricted ISI distributions simulated in this study.

Although the deconvolution technique exhibited significant advantages compared to event-related averaging, these simulation results may only be directly applied to situations in which there is a constant response to the first of the two sequentially dependent stimuli. For instance, the duration of the response function to event type A(1) was independent of the temporal interval between A(1) and B(3). Due to this independence, a deconvolution model that estimates the 20 timepoints following event onset adequately accounted for the variance in the BOLD signal induced by the response to A(1). However, if the response profile to event A(1) increased in duration as the ISI increased, then modeling a constant number of timepoints following A(1) would inadequately account for the sustained profile of the response function. While this situation complicates the modeling process, it does not change the general conclusions of these simulations; if the response to event A(1) was sustained across the ISI, the distortions evident in the event-related averages would still remain.

### *Compound versus partial trial designs*

If the design of a compound trial experiment prohibits the use of temporal jitter between successive stimulus presentations, then using a partial trial design is the only way to accurately reconstruct the event-related BOLD signal (Ollinger et al., 2001a, 2001b). However, if a temporally jittered ISI is used, then deconvolution of the evoked BOLD response is possible using both compound and partial trial methods. The present simulations suggest that the partial trial method produced more robust estimates of the event-related timeseries compared to the compound trial method when the standard deviation of the ISI distribution was small. As the standard deviation of the ISI distribution increased, the compound trial method produced results that were comparable to the partial trial method. Thus, from a purely statistical standpoint, the partial trial approach generally produces more reliable estimates of the BOLD response.

While the partial trial approach compares favorably to the compound trial approach on statistical grounds, the assumptions underlying the method are potentially problematic. The effective use of partial trials hinges on the assumption that the response to the first of two sequential stimuli is unaffected by the presence or the absence of the second stimulus. The specific concern is that subjects may alter their overarching cognitive set during the experiment if they have knowledge about the probability of a particular stimulus event occurring. For instance, there may by important cognitive consequences of a subject knowing that 30% of the test stimuli will not be presented in a delayed match-to-sample task. Ideally, subjects would encode and rehearse the sample stimulus with equivalent vigilance on all trials. However, subjects may engage in probability matching, remembering the sample stimulus on a random set of 70% of the trials and not performing the task at all on 30% of the trials (e.g., Jonides, 1983). Another possibility is that subjects may allocate effort to the task in a graded fashion that corresponds to the probability of receiving a test stimulus (e.g., Johnson and Yantis, 1995). If subjects use the former strategy, and perform the task only on a proportion of the trials, then serious problems may arise in the interpretation of the event-related BOLD timeseries estimated using the partial trial approach. However, if subjects adopt the second strategy, then modulating their overall task engagement in accordance with the probability of a partial trial would presumably result in a reduction of experimental power.

Fortunately, experimenters may determine the extent to which subjects adopt these deviant strategies by looking at the behavioral responses collected in the scanner or in pilot experiments outside of the scanner. For instance, the probability matching hypothesis predicts a mixture-distribution of reaction times (Johnson and Yantis, 1995; Yantis et al., 1991). In contrast, the graded allocation account predicts a uniform RT distribution, but main effects of reaction time when the number of partial trials is parametrically varied. If the results do suggest that subjects are complying with task demands, then partial trials are ideal for achieving good estimation and statistical power, while at the same time using a restricted range of ISIs that may allow tighter control over the psychological validity of the experimental paradigm.

## Conclusions

Choosing a method for computing event-related averages in fMRI depends on the relative weight that a researcher places on several competing factors. First, psychological validity and the ability to detect differences between conditions dictate that short ISIs be used. Short ISIs minimize the time that the subject's behavior can vary unpredictably and maximize the number of trials that can be presented in a given unit of time. However, the ability to correctly estimate the shape of the event-related BOLD response diminishes as the range of the ISI distribution decreases. This loss of estimation power is undesirable because accurately characterizing cognitive functions requires a consideration of the temporal characteristics of the BOLD response in a given brain region.

The results of the present simulations show that when the order of stimulus presentation is randomized, event-related averaging and deconvolution techniques produce comparably robust results in terms of statistical power. Parsimony would dictate that using event-related averages may be preferable in this case; computing event-related averages is relatively straightforward and complicated baseline issues are largely nullified by randomized stimulus ordering. In addition, event-related averages allow a characterization of the BOLD timecourse before time '0' (or the time of event onset). This information can be useful when evaluating preparatory activity before an event occurs, and is not easily obtained using modeling techniques such as deconvolution (Connolly et al., 2002; Yantis et al., 2002). However, when there is a complete dependence in the ordering of stimulus events, event-related averaging compares unfavorably to deconvolution techniques, even when partial trials are used. The distortions in the timeseries are so large that very different conclusions would be drawn based on the event-related average timecourses and the deconvolution timecourses

(e.g., compare Figs. 5a, g or 7a–d and g–l). Finally, the partial trial technique is promising for maximizing statistical power and estimation efficiency, especially when restricted ISI distributions are used. If introducing partial trials does not affect subject's cognitive set, then the partial trial technique is also ideal from a psychological standpoint in that it allows most of the events to be presented closer together in time.

## Acknowledgments

Thanks to Amy L. Shelton and Steven Yantis, who provided valuable feedback and resources. Thanks also to Joseph B. Sala and Edward Awh for useful discussions, and two anonymous reviewers for insightful suggestions. Supported by a National Science Foundation Graduate Research Fellowship.

## References

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

Bandettini, P.A., 1993. MRI studies of brain activation: dynamic characteristics. Functional MRI of the brain. Soc. Magn. Reson. Med., 144–151 (Berkeley).

Botvinick, M., Nystrom, L.E., Fissell, K., Carter, C.S., Cohen, J.D., 1999. Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature 402, 179–181.

Boynton, G.M., Engel, S.A., Glover, G.H., 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., Petersen, S.E., Raichle, M.E., 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. U. S. A. 93, 14878–14883.

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

Buracas, G.T., Boynton, G.M., 2002. Efficient design of event-related fMRI experiments using M-sequences. NeuroImage 16, 801–813.

Burock, M.A., Dale, A.M., 2000. Estimation and detection of event-related fMRI signals with temporally correlated noise: a statistically efficient and unbiased approach. Hum. Brain Mapp. 11, 249–260.

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, 3735–3739.

Cohen, M.S., 1997. Parametric analysis of fMRI data using linear systems methods. NeuroImage 6, 93–103.

Connolly, J.D., Goodale, M.A., Menon, R.S., Munoz, D.P., 2002. Human fMRI evidence for the neural correlates of preparatory set. Nat. Neurosci. 5, 1345–1352.

Corbetta, M., Kincade, J.M., Ollinger, J.M., McAvoy, M.P., Shulman, G.L., 2000. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat. Neurosci. 3, 292–297.

Corbetta, M., Kincade, J.M., Shulman, G.L., 2002. Neural systems for visual orienting and their relationships to spatial working memory. J. Cogn. Neurosci. 14, 508–523.

Courtney, S.M., Ungerleider, L.G., Keil, K., Haxby, J.V., 1997. Transient and sustained activity in a distributed neural system for human working memory. Nature 386, 608–611.

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

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

Fransson, P., Kruger, G., Merboldt, K.D., Frahm, J., 1999. Temporal and spatial MRI responses to subsecond visual activation. Magn. Reson. Imaging 17, 1–7.

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

Friston, K.J., Holmes, A., Worsley, K., Poline, J., 1995b. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

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

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. 12(Suppl 2), 76–89.

Hopfinger, J.B., Buonocore, M.H., Mangun, G.R., 2000. The neural mechanisms of top-down attentional control. Nat. Neurosci. 3, 284–291.

Johnson, D.N., Yantis, S., 1995. Allocating visual attention: tests of a two-process model. J. Exp. Psychol. Hum. Percept. Perform. 21, 1376–1390.

Jonides, J., 1983. Further toward a model of the mind's eye's movements. Bull. Psychonom. Soc. 21, 247–250.

Kirino, E., Belger, A., Goldman-Rakic, P., McCarthy, G., 2000. Prefrontal activation evoked by infrequent target and novel stimuli in a visual target detection task: an event-related functional magnetic resonance imaging study. J. Neurosci. 20, 6612–6618.

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, 759–773.

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, 735–759.

Ollinger, J.M., Corbetta, M., Shulman, G.L., 2001a. Separating processes within a trial in event-related functional MRI. NeuroImage 13, 218–229.

Ollinger, J.M., Shulman, G.L., Corbetta, M., 2001b. Separating processes within a trial in event-related functional MRI. NeuroImage 13, 210–217.

Sala, J.B., Rama, P., Courtney, S.M., 2003. Functional topography of a distributed neural system for spatial and nonspatial information maintenance in working memory. Neuropsychologia 41, 341–356.

Stark, C.E., Squire, L.R., 2001. When zero is not zero: the problem of ambiguous baseline conditions in fMRI. Proc. Natl. Acad. Sci. U. S. A. 98, 12760–12766.

Wager, T.D., Nichols, T.E., 2003. Optimization of experimental design in fMRI: a general framework using a genetic algorithm. NeuroImage 18, 293–309.

Woldorff, M.G., 1993. Distortion of ERP averages due to overlap from temporally adjacent ERPs: analysis and correction. Psychophysiology 30, 98–119.

Yantis, S., Meyer, D.E., Smith, J.E., 1991. Analyses of multinomial mixture distributions: new tests for stochastic models of cognition and action. Psychol. Bull. 110, 350–374.

Yantis, S., Schwarzbach, J., Serences, J.T., Carlson, R.L., Steinmetz, M.A., Pekar, J.J., Courtney, S.M., 2002. Transient neural activity in human parietal cortex during spatial attention shifts. Nat. Neurosci. 5, 995–1002.