NeuroImage 60 (2012) 1788–1799

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Full Length Article

Optimization of contrast detection power with probabilistic behavioral information

Dietmar Cordes [a,*], Grit Herzmann [b], Rajesh Nandy [c], Tim Curran [b]

[a] Department of Radiology, School of Medicine, University of Colorado-Denver, CO 80045, USA
[b] Department of Psychology and Neuroscience, University of Colorado-Boulder, CO 80309, USA
[c] Departments of Biostatistics and Psychology, UCLA, Los Angeles, CA 90095, USA

ARTICLE INFO

Article history:
Received 3 June 2011
Revised 14 January 2012
Accepted 28 January 2012
Available online 6 February 2012

Keywords:
fMRI
Genetic algorithm
Event optimization
Familiarity
Recollection
Recognition memory

ABSTRACT

Recent progress in the experimental design for event-related fMRI experiments made it possible to find the optimal stimulus sequence for maximum contrast detection power using a genetic algorithm. In this study, a novel algorithm is proposed for optimization of contrast detection power by including probabilistic behavioral information, based on pilot data, in the genetic algorithm. As a particular application, a recognition memory task is studied and the design matrix optimized for contrasts involving the familiarity of individual items (pictures of objects) and the recollection of qualitative information associated with the items (left/right orientation). Optimization of contrast efficiency is a complicated issue whenever subjects' responses are not deterministic but probabilistic. Contrast efficiencies are not predictable unless behavioral responses are included in the design optimization. However, available software for design optimization does not include options for probabilistic behavioral constraints. If the anticipated behavioral responses are included in the optimization algorithm, the design is optimal for the assumed behavioral responses, and the resulting contrast efficiency is greater than what either a block design or a random design can achieve. Furthermore, improvements of contrast detection power depend strongly on the behavioral probabilities, the perceived randomness, and the contrast of interest. The present genetic algorithm can be applied to any case in which fMRI contrasts are dependent on probabilistic responses that can be estimated from pilot data.

Introduction

Optimal design efficiency refers to the best arrangement of stimuli in both event-related and block-type fMRI tasks to make tasks more efficient by reducing scanner time and saving costs. Unfortunately, there is no universally optimum arrangement of events and the best arrangement of stimuli is strongly dependent on whether activation is to be detected, a specific contrast is to be obtained, or if the hemodynamic response function (HRF) is to be determined (Dale, 1999; Friston et al., 1999; Josephs and Henson, 1999). When determining the best shape of the hemodynamic response, it is customary to describe this scenario as optimization of estimation efficiency, whereas best detection of activation or a contrast of activation is referred to as optimization of detection power. It is well known that, in general, optimized random designs are best to determine the shape of the HRF, whereas block-type designs are best for detection power (at least for simple contrasts involving the difference of activation patterns) (Birn et al., 2002; Friston et al., 1999). The converse is also true: block-type designs are inefficient to determine the HRF, and random designs are inefficient to detect activation (Liu et al., 2001).

Other factors in the design of events are psychological requirements to avoid habituation and anticipation effects, which can be achieved by counterbalancing stimuli to some degree. Minimizing predictability of events leads to more random designs that decrease the detection power. To maximize detection power, it is necessary to relax the condition of stimulus randomness and only require a "perceived" randomness of stimuli that is determined to be sufficient based on pilot data (Liu, 2004; Liu and Frank, 2004; Liu et al., 2001).

Design optimization and anticipated behavioral responses

Of particular interest in psychology is the design of tasks for fMRI in which analysis depends on specific probabilistic behavioral outcomes such as response accuracy and latency. The design matrix is set up by sorting the fMRI trial responses post-hoc according to behavioral measurements. Predicting the optimal arrangement of events for maximum contrast detection power requires optimization that must include the behavioral probabilities of the anticipated responses. Anticipated behavioral constraints have not been studied in the optimization of contrast detection power or design efficiency, and available software for design optimization cannot handle them.

* Corresponding author at: Univ. of Colorado Denver, Dept. of Radiology, School of Medicine, 12700 E. 19th Ave., C278, RC2/P15-1206, Aurora, CO 80045, USA. Fax: +1 866 372 2720.
E-mail address: dietmar.cordes@UCDenver.edu (D. Cordes).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2012.01.127

Several natural questions, which are the focus of the present research, arise:

1. Can the order of the stimulus sequence be optimized for maximum detection power based on the probability of the behavioral outcomes from previous pilot studies?
2. Neglecting predictability of stimuli, is a block-type design the most efficient design for contrast detection?
3. How good is a random design for contrast detection? What are the tradeoffs between detection power and perceived randomness in this case?
4. How robust is an optimized design based on probabilistic behavioral information if the anticipated behavioral outcome is less accurate due to the fact that subjects' accuracy and timing may have improved?

To answer these questions, we newly designed a genetic algorithm that allows the incorporation of probabilistic behavioral information and optimized the design for maximum contrast detection power. Simulations were carried out for a recognition memory task in which responses are highly probabilistic, making this task an ideal test for the proposed genetic algorithm. Functional MRI data for a group of 18 subjects were obtained with an experimental design that was optimized by a new algorithm. Results were compared to findings in the literature.

Recognition memory: familiarity and recollection

The critical role of the hippocampus and nearby medial temporal lobe (MTL) cortex in learning and memory is well documented, especially from neuropsychological studies of anterograde amnesia that results from damage to these areas (Eichenbaum et al., 2007; Squire et al., 2004). As research in this domain has progressed, theories of the precise contributions of these regions to learning and memory have increasingly called for differentiation between the hippocampus proper and surrounding MTL cortex. For example, behavioral and electrophysiological work has suggested that recognition memory is supported by separate processes: those that support the *recollection* of the details of previous experiences and those that allow us to recognize events based on *familiarity* without the recall of specific details (Rugg and Curran, 2007; Yonelinas, 2002). Some researchers have hypothesized that recollection is specifically related to the hippocampus and parahippocampal cortex, whereas familiarity is related to perirhinal cortex (Aggleton and Brown, 1999; Norman and O'Reilly, 2003). Others have suggested that the hippocampus and MTL cortex work together in an undifferentiated manner to support both types of memory (Squire et al., 2007; Wixted and Squire, 2011). Initial tests of these ideas in humans have focused on patients with selective hippocampal damage, but have been impeded by inconsistent results (Squire et al., 2007; Yonelinas et al., 2010). More recent work has tried to address these issues with fMRI (reviewed by Carr et al., 2010), an enterprise that demands the development of advanced high-resolution imaging techniques to resolve differences between hippocampus and MTL regions.

In the present experiment, recollection was defined as the recognition of the study orientation for common objects whereas familiarity was defined as recognition without the recollection of orientation. Subjects studied pictures of objects followed by recognition memory tests that required recollecting the original left/right orientation of each studied picture. Functional MRI scanning took place during recognition testing. Test lists contained studied pictures in their original ("same") orientation, studied pictures in the opposite ("different") orientation, and non-studied ("new") pictures. Activation related to recollection was estimated by contrasting trials in which subjects correctly classified the orientation of studied items as "same" or "different," with trials in which subjects incorrectly classified the study orientation of old items. Here we assume that these two sets of trials differ with regard to subjects' recollection of picture orientation. Activation related to familiarity was estimated by contrasting trials in which subjects incorrectly classified the study orientation of items (but recognized that the items were old) with trials in which subjects correctly classified non-studied pictures as "new." Here we assume that these two sets of trials differ with regard to the familiarity of old vs. new items, with minimal contributions of recollection because orientation was judged incorrectly. The logic of these contrasts is similar to past fMRI research using source recognition to separate recollection and familiarity (Diana et al., 2007; Spaniol et al., 2009). Similar to typical source recognition tasks, there are at least two limitations to keep in mind when interpreting results. First, recollection and familiarity contrasts are likely to also differ with regard to confidence (Wixted and Squire, 2011). Second, the familiarity contrast may include activity related to the recollection of attributes other than orientation, so-called "noncriterial" recollection (Parks, 2007; Yonelinas and Jacoby, 1996). Although these issues are critical for psychological interpretation, they are less important for our primary present goal of design optimization.

Methodology

In this section we briefly review the parameterization of the HRF, the form of the general linear model and its solution in the presence of temporal autocorrelations, and the formal definition of design efficiency and contrast detection power. Furthermore, we define a nonpredictability index to avoid psychological confounds and introduce a measure to determine the robustness of the design against misspecification of behavioral information.

Contrast detection power and the linear model

According to the general linear model (GLM), the relationship between stimuli and the BOLD response is modeled as a convolution of stimulus functions $s_r(t)$with amplitude β_r and the HRF $h(t)$ which we assume to be known for the purpose of this research. In particular, we assume that the HRF has the conventional two-gamma form

$$h(t) = \left(\frac{t}{d_1}\right)^{a_1} e^{\frac{-(t-d_1)}{b_1}} - c_1\left(\frac{t}{d_2}\right)^{a_2} e^{\frac{-(t-d_2)}{b_2}}$$

with parameters $a_1 = 6,\ a_2 = 16,\ b_1 = 1,\ b_2 = 1,\ c_1 = \frac{1}{6},\ d_1 = a_1 b_1,$ and $d_2 = a_2 b_2$ similar to Glover (1999).

The fMRI signal $y(t)$ is then given by

$$y(t) = \sum_{r=1}^{R} \beta_r s_r(t) * h(t) + \varepsilon(t) \text{ for } t = 1, \ldots, T \tag{1}$$

where R is the number of stimulus functions and $\varepsilon(t)$ is a Gaussian distributed error term of the form $N(0, \Sigma)$ with mean zero and autoregressive (AR) order 1 such that the elements of the covariance matrix are given by

$$\Sigma'_{lm} = \frac{\sigma^2}{1-\phi^2}\phi^{|l-m|} \tag{2}$$

where σ^2 is the variance and ϕ is the autocorrelation coefficient (Cordes and Nandy, 2007). In matrix notation, Eq. (1) becomes

$$y = X\beta + \varepsilon \tag{3}$$

where y is a column vector corresponding to the observed signal at a particular voxel and X is the $T \times R$ design matrix resulting from the convolution of

$$\beta_r s_r(t) * h(t), \quad r = 1, \ldots, R,$$

sampled at the TR. To obtain uncorrelated errors, standard pre-whitening is performed by multiplying Eq. (3) with the matrix K such that

$$K\Sigma K' = \sigma^2 I,$$

where the prime indicates transpose and I labels the unit matrix (Friston et al., 2000). Low-frequency drifts can be projected out by applying a high-pass filter with cutoff frequency f_0. This is accomplished by transforming Eq. (1) into frequency space using the Fourier transform and setting all frequencies f for $|f|<f_0$ to zero (ideal high-pass filter). This operation will transform the variables in Eq. (3). Note that high-pass filtering and pre-whitening are commutative operations. In the following to simplify notation, we assume that high-pass filtering (see for example Gonzalez and Woods, 1993) and pre-whitening (see for example Wager and Nichols, 2003) have been carried out on Eq. (3). Then, given the transformed data y and the transformed design matrix X (which is different from the y and X in Eq. (3)), the least squares solution of β is given by

$$\beta_{LS} = \left(X'X\right)^{-1}X'y$$

with variance–covariance matrix var $\beta_{LS} = \sigma^2(X'X)^{-1}$.

In these equations, we have used the same symbols as before in Eq. (3) to simplify notation. However, the reader should keep in mind that now X and y refer to the transformed (preprocessed) variables.

For a given contrast matrix $C = [c_1 c_2 \ldots c_p]'$ with p contrast vectors, the variance of the least square estimate $C\beta_{LS}$ becomes

$$\mathrm{var}(C\beta_{LS}) = \sigma^2 C\left(X'X\right)^{-1}C'$$

and the contrast detection power is defined as

$$\xi = \frac{1}{trace[diag(w)\,\mathrm{var}(C\beta_{LS})]} = \frac{1}{\sigma^2 trace\left[diag(w)C\left(X'X\right)^{-1}C'\right]}$$

where w is a suitable weight vector describing the importance of each contrast vector c_i for $i=1,\ldots,p$, and $diag(w)$ a diagonal matrix with the elements of w in its diagonal. This equation can also be used to compute the design estimation efficiency to determine the shape of the HRF if finite impulse response functions are used as a basis set to model the HRF. Note that, strictly speaking, ξ depends also on the error variance σ^2, as pointed out by Mechelli et al. (2003). In the current research, we neglect this dependency and treat σ^2 as a constant by setting it to 1.

Counter balancing of stimuli order to reduce the chance of prediction

Predictability, in general, refers to the correct guessing of a future event based on the memory of a sequence of similar past events. A block design is highly predictable, because the same types of stimuli are presented, and a completely random design is non-predictable. Given a sequence of events (stimuli) {ijk…}, we define an index of "non-predictability" or randomness in the range [0,1], where 1 means the next event is perfectly non-predictable and 0 means it is perfectly predictable. Thus, predictability needs to be calculated based on how many stimuli that were presented previously may influence the response of the next stimulus to be presented. For the present recognition paradigm, we have three different stimuli ("same," "different," "new"). If a "same" stimulus was presented, then the next stimulus to be presented, in order to be non-predictable, must be either "same," "different," or "new" with equal probability of 1/3. This is called first-order non-predictability. Higher order non-predictability takes into account more than just the last stimulus/response presented to predict the next stimulus. We defined the predictability of order one to three as follows:

1. order: Let p_i be the probability that the ith stimulus (example $i=1, 2, 3$) occurs next, independent of the previous stimulus. Thus, if there are n different stimuli (example $n=3$), the stimuli are perfectly balanced if $p=1/n$ for all i, and the non-predictability index of first order, $I(p_i)$, is equal to 1.
2. order: Let $p_{j|i}$ be the probability that whenever stimulus i occurred, the next presented stimulus is stimulus j. Also here, the design is perfectly balanced if $p_{j|i}=1/n$ for all i,j. Then, the non-predictability index of second order, is $I(p_{j|i})=1$.
3. order: Let $p_{k|ij}$ be the probability that whenever stimulus i occurred and the next stimulus was stimulus j, then the next presented stimulus is stimulus k. Also here, the design is perfectly balanced if $p_{k|ij}=1/n$ for all i,j,k. Then, the non-predictability index of order three is $I(p_{k|ij})=1$.

For all other values of $p_i, p_{j|i}, p_{k|ij}$ we define the non-predictability indices $I(p_i), I(p_{j|i}), I(p_{k|ij})$ as linearly scaled functions of $1-\max_\theta|p_\theta - E(p_\theta)|$ mapped to the interval [0,1] where p_θ is either $p_i, p_{j|i}$, or $p_{k|ij}$ and $E(p_\theta)$ is the expectation value for a perfectly balanced design, i.e.

$$I(p_\theta) = 1 - \frac{\max_\theta|p_\theta - E(p_\theta)|}{1-E(p_\theta)}.$$

Please note that the specified non-predictability indices are similar to the ones defined by Wager and Nichols (2003), however, in our definition we use conditional probabilities and chose a different meaning for the order of the non-predictability indices. A main difference of the defined criteria for non-predictability is that other criteria (Kao et al., 2009; Wager and Nichols, 2003) do not assume that the different stimulus types are happening equally often. The specified non-predictability constraints are implemented as hard constraints in this research. Hard constraints have the advantage that specified criteria are exactly met whereas soft constraints are only met up to a specified probability or soft threshold yielding an overall solution of the constrained optimization problem. In this research, we chose to use hard constraints because we wanted to exactly meet the specified degrees for non-predictability. In general, the advantage of soft constraints over hard constraints is that faster algorithms can be found leading to a solution of the optimization problem. Furthermore, soft constraints can lead to increased optimal values because the space of constraints has more degrees of freedoms available than the space of hard constraints.

For completeness, we would like to mention that other criteria based on soft constraints have been proposed previously. For example, Wager and Nichols (2003) and Kao et al. (2009) proposed a multi-objective criterion. In particular, the criterion proposed by Kao et al. (2009) is a multi-objective optimal experimental design which is an improvement of the weighted average design criteria of Wager and Nichols (2003). While the algorithm of Kao et al. seems to perform better than the algorithm of Wager and Nichols, this better performance is not related to using soft or hard constraints of the multi-objective design criterion.

Robustness against misspecification of behavioral information

To measure the robustness of the obtained design against misspecification of the probabilities of the behavioral information, we calculated the mean ratio of the contrast detection power by

$$\xi_{1|2} = \underset{\text{subjects}}{\text{mean}} \frac{\xi(\text{misspecified design})}{\xi(\text{optimal design})}$$

where the misspecified design of a subject differs in the behavioral probabilities based on actual behavioral subject data collected during fMRI scanning. Here, the optimal detection power, ξ (optimal design), is a theoretical quantity that is obtained by optimizing the design for the actual achieved behavioral probabilities during fMRI scanning.

Materials and methods

Subjects

Subjects were 18 healthy undergraduate students from the University of Colorado at Boulder: 10 female, 8 male, mean age 21.9 years, SD = 3.03, all right-handed. Subjects had previously completed (approximately a week earlier) the same experiment with EEG recording, but using different pictures as stimuli. The findings of the EEG experiment and the EEG–fMRI relationships derived from both the EEG and present experiment are reported elsewhere (Herzmann et al., submitted for publication).

fMRI acquisition

fMRI was performed in a 3.0 T GE HDx MRI scanner equipped with an 8-channel head coil and parallel imaging acquisition using EPI with imaging parameters: ASSET = 2, ramp sampling, TR/TE = 1.5 s/30 ms, FA = 70°, FOV = 22 cm × 22 cm, thickness/gap = 3.5 mm/0.5 mm, 30 slices, resolution 64 × 64, axial acquisitions. A standard 2D co-planar T1-weighted image and a standard 3D high resolution T1-weighted SPGR (1 mm^3 resolution) were also collected.

Memory task

Subjects studied a long list of 268 pictures of common objects that were asymmetric about the vertical axis. Functional scanning took place a day after the study session during memory testing with lists that contained pictures studied in the subjects' original orientation, pictures studied in the opposite left/right orientation, and new pictures never studied. The length of presentation of each stimulus was 3 s. During scanning, 402 pictures (about 134 same orientation, 134 different orientation, and 134 new pictures) were presented. The actual number of stimuli presented varied from 134 by less than 2.5% due to practical implementations of the genetic algorithm. Subjects responded within the 3-second time period of stimulus presentation by selecting one of three memory judgments for each stimulus: studied picture with "same" orientation, studied picture with "different" orientation, or "new". The conditions were coded according to the stimulus (second letter) and the subject's response (first letter):

1. s|s (stimulus is "same", subject responds "same")
2. d|d (stimulus is "different", subject responds "different")
3. d|s (stimulus is "same", subject responds "different")
4. s|d (stimulus is "different", subject responds "same")
5. n|n (stimulus is "new", subject responds "new").

In the data modeling, we disregarded the other four possible responses (n|s (stimulus is "same", subject responds "new"), n|d (stimulus is "different", subject responds "new"), s|n (stimulus is "new", subject responds "same"), and d|n (stimulus is "new", subject responds "different")) because these scenarios are not relevant for the contrast of interest and occurred only with very low probability, with the exception of a single subject (#11 in Table 1), according to our pilot studies.

In order to maximize contrast detection power, the stimulation periods were not interleaved by null events (such as resting periods) and the pictures were presented one after the other. Such an arrangement could potentially result in some contributions from the nonlinear BOLD effect. However, the nonlinear effect is unknown for memory activation and so far has only been systematically investigated for primary motor and visual sensory cortex (Wager et al., 2003). For motor and visual cortex, the nonlinear BOLD response has been found significant only for stimulation periods of less than 2 s (Buckner, 1998). Consequently, an investigation of the existence of nonlinear effects for memory activation at a stimulation period of 3 s is beyond the scope of the current research study.

Table 1
Actual behavioral probabilities during fMRI scanning.

Subject #	p(s|s)	p(d|s)	p(s|d)	p(d|d)	p(n|n)
1	0.89	0.11	0.41	0.59	0.94
2	0.72	0.28	0.40	0.60	0.92
3	0.92	0.08	0.23	0.77	0.97
4	0.80	0.20	0.35	0.65	0.81
5	0.93	0.07	0.27	0.73	0.98
6	0.83	0.17	0.15	0.85	0.93
7	0.89	0.11	0.34	0.66	0.89
8	0.85	0.15	0.23	0.77	0.95
9	0.93	0.07	0.36	0.64	0.94
10	0.73	0.27	0.31	0.69	0.99
11	0.84	0.16	0.48	0.52	0.64
12	0.96	0.04	0.23	0.77	0.99
13	0.89	0.11	0.29	0.71	0.96
14	0.95	0.05	0.22	0.78	0.94
15	0.93	0.07	0.15	0.85	0.98
16	0.98	0.02	0.13	0.87	0.98
17	0.94	0.06	0.15	0.85	0.98
18	0.90	0.10	0.34	0.66	0.98
Mean	0.88	0.12	0.28	0.72	0.93
std	0.08	0.08	0.10	0.10	0.09
*Mean pilot study	0.78	0.11	0.27	0.60	0.87

Note: The five conditions are s|s ("same" stimulus, subject responds "same"), d|s ("same" stimulus, subject responds "different"), s|d ("different" stimulus, subject responds "same"), d|d ("different" stimulus, subject responds "different"), n|n ("new" stimulus, subject responds "new"). The last line (indicated by *) gives the probabilities obtained from pilot studies.

Genetic algorithm implementation

Genetic algorithms are suitable to solve optimization problems in a high dimensional space (Ahn, 2006). Starting from a large number of randomly generated design vectors, where the sequence of stimuli is coded using discrete numbers for different conditions, three processes act on the vectors in genetic algorithms: selection, crossover, and point mutations.

In addition, specialized new designs such as random designs, block designs and their combinations can be added to the population vector. The process of adding specialized new designs is referred to as "immigration" (Kao et al., 2009). The purpose of immigration is to add extra variability, which prevents the genetic algorithm from being trapped in a local optimal solution.

The selection process uses particular fitness criteria (in our case contrast detection power) to select the best vector of the particular generation. Crossover interchanges a portion of the vector sequence between a pair of vectors using a randomly selected cut point, and point mutation changes a proportion of the entries of a vector sequence into different ones. For fMRI, the first algorithm for design optimization using a genetic algorithm was proposed by Wager and Nichols (2003). For our research, we designed an algorithm similar to Wager and Nichols and incorporated changes so that probabilistic behavioral constraints can be included in the optimization.

Specifically, for the proposed memory task the matrix of temporal autocorrelations is set up with $\phi = 0.2$ in Eq. (2) (Cordes and Nandy, 2007). Then, the first generation is defined by 500 randomly generated design vectors (also called population vectors). Each design vector has the entry 1, 2, or 3 specifying which stimulus was shown ("1" for "same," "2" for "different," "3" for "new"). Each entry in the design vector corresponds to a stimulus duration of 3 s. Then, the three stimulus vectors representing the timing of stimuli 1, 2, and 3,

respectively, are computed for the entire population. Using the predetermined behavioral information from pilot data, the probabilities p(s|s), p(d|s), p(s|d), p(d|d), p(n|n) are used to construct the timing of the five regressor delta functions. Upon convolution with the assumed hemodynamic response function, following pre-whitening and high pass filtering, the 5-column design matrix is formed and the contrast detection power can be calculated for the entire population. Since the behavioral data are probabilistic, the median detection power is computed among 100 repetitions. This completes the probabilistic loop. The next step is to sort the population vectors in descending order according to the median value of contrast detection power. The best 25 population vectors (#1 to #25) are used to perform a random pair-wise crossover. The final number of new vectors arising from the crossover operations is 450. A new population of design vectors (next generation) is then defined by specifying #1 to #11 equal to the best vector of the previous population (best vector #1 + 10 replications of best vector #1) and #12 to #461 as the 450 crossover vectors from the previous step. Then, a point-wise mutation with probability 0.01 is carried out on all vectors of the new generation except vector #1. Point-wise mutation with probability 0.01 here means that the "stimulus" numbers for the "same", "different" and "new" stimuli are randomly assigned for 1% of the entries of each vector. Finally, random vectors are added to the new generation such that the total size of the generation contains 500 population vectors. The program then repeats until convergence or until a fixed number of generations are computed.

For this research project all simulations were run initially up to generation 1000. From generation 100 to generation 1000, we obtained the empirical result that the median detection power did not increase by more than 1%. From generation 200 to generation 1000, we could not detect reliably any increase due to the small intrinsic fluctuations of the probabilistic loop, and the median detection power was essentially flat. Generation 100 is an acceptable threshold to achieve convergence and was used to stop the optimization process. It is also possible to define convergence by calculating

$$\left|\frac{\xi(g+1)-\xi(g)}{\sigma(\xi(g))}\right| < \varepsilon$$

where $\xi(g+1)-\xi(g)$ is the difference of the median detection power at generation $g+1$ and g, $\sigma(\xi(g))$ is the standard deviation of the median detection power at generation g, and ε is a convenient threshold (such as 0.001). However, such an approach is more computationally expensive due to the estimation of $\sigma(\xi(g))$ and was not carried out due to time constraints.

The genetic variables involved were optimized by varying the number of random lists, the number of crossover vectors to be used in the crossover computations, the number of cross-over computations performed, and the number of replications of the best vector. The numbers stated above lead to the fastest performance. A flowchart of the entire algorithm is given in Fig.1.

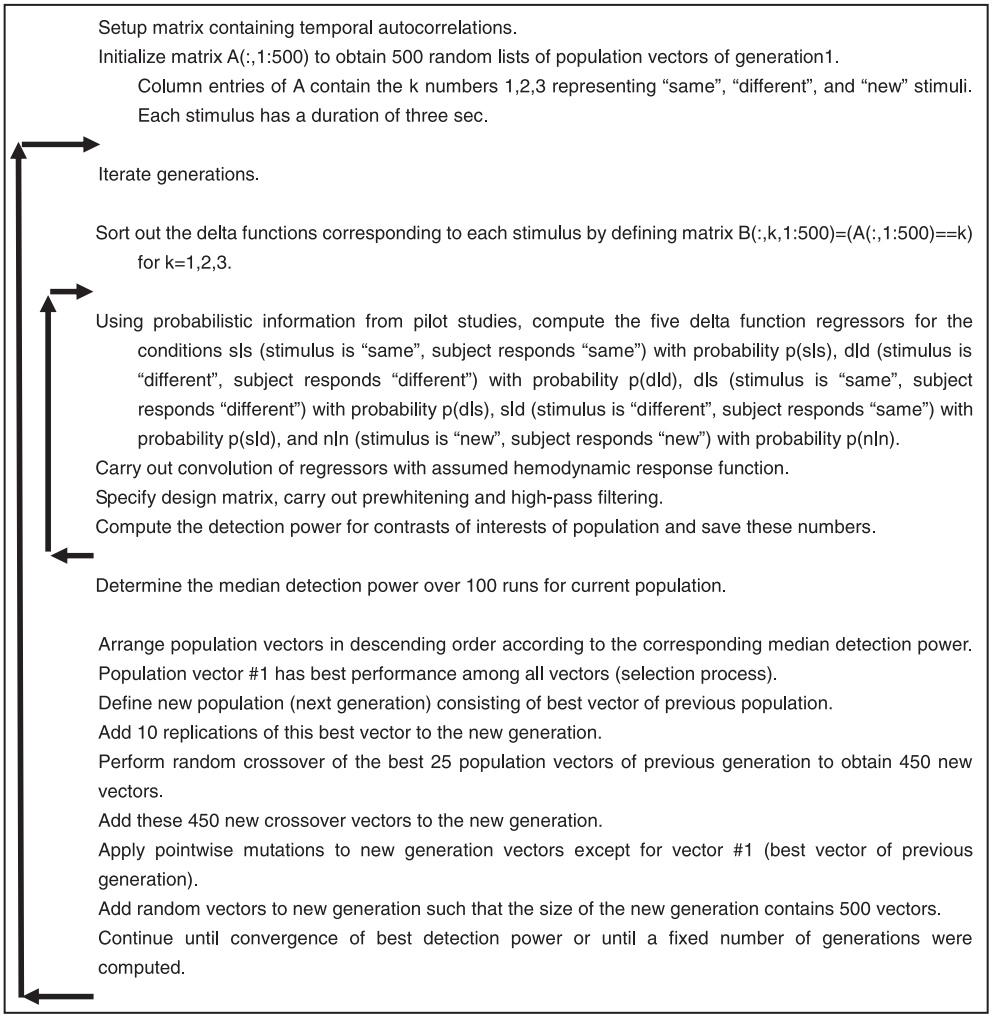

Fig. 1. Flowchart of the genetic algorithm. The probabilistic information of the behavioral responses is included in the optimization by an additional loop (inner loop).

To incorporate the non-predictability of the stimulus sequence, if desired, all vectors determined for the next generation are subjected to the criteria for non-predictability, and the operations using cross-over, point-wise mutations, and random-vector-adding are repeated until the criteria for non-predictability are satisfied and the necessary number of 500 new vectors for the next generation is found.

We would like to point out that the proposed algorithm under the condition that no behavioral constraints are included or all behavioral information has a probability of one leads to deterministic regressors of the design matrix. In this case, our algorithm is fundamentally equivalent to the method proposed by Wager and Nichols (2003).

The fMRI task was designed as described before in section "Memory task" and was programmed in EPRIME (Psychology Software Tools, INC., Pittsburgh, PA). The behavioral probabilities used in the design were determined from pilot data from the same design with an independent sample of undergraduate subjects. The average probabilities were p(s|s) = 0.78, p(d|s) = 0.11, p(s|d) = 0.27, p(d|d) = 0.6, and p(n|n) = 0.87. The actual probabilities during fMRI scanning were slightly different and are listed in Table 1. Note that, in general, subjects were more likely to respond "same" than "different" (i.e. p(s|s) > p(d|d) and p(s|d) > p(d|s)) indicating a bias for "same." Using the proposed genetic algorithm, the fMRI task was optimized using the average probabilities from the pilot study. Sufficient balance and randomness were accounted for by specifying non-predictability indices = {0.975, 0.9, 0.85} for 1st to 3rd order. Due to GE hardware limitations of the MR scanner prohibiting EPI scans with more than 20,000 image acquisitions, the task was split into two runs of 201 stimuli each. Scan duration for each task was 603 s after 10 s of equilibrium null scans at the beginning. Behavioral data were collected using a conventional 4-button response box with EPRIME.

The subjects' expected accuracy was incorporated in the optimization of the design by an additional loop (inner loop) as shown in Fig. 1, which outlines the structure of the proposed genetic algorithm. The inner loop uses the probabilistic information of the behavioral responses to extract the possible five regressors for the conditions s|s, d|s, s|d, d|d, and n|n. These regressors are then convolved with the two-gamma HRF giving the design matrix. Then, contrast efficiency for recollection and familiarity are computed using equal weighting.

Data analysis

All fMRI data were realigned in SPM5 (http://www.fil.ion.ucl.ac.uk/spm/) and corrected for differences in timing of slice acquisitions. The design matrix was set up using the five regressors for conditions s|s, d|s, s|d, d|d, and n|n, which were formed by using the stimulus sequence together with the collected behavioral information, as described before. The design matrix and all voxel time series were high-pass filtered using the standard cut-off frequency 1/120 Hz (Frackowiak, 2004). No temporal low-pass filtering was carried out. A brain mask was used to effectively eliminate all non-brain voxels leading to an average of about 1200 voxels per slice. Standard smoothing using a Gaussian FWHM = 6 mm was carried out to increase the SNR and to enable group analysis. The data were re-sliced to an isotropic voxel size of $2\times2\times2$ mm^3. Two contrasts of fMRI memory activation – recollection and familiarity – were computed according to past fMRI research (Diana et al., 2007; Spaniol et al., 2009). These contrasts assume that recollection is indexed by the ability to correctly remember item and orientation (i.e., source), whereas familiarity is indexed by item recognition without the recollection of orientation (i.e., source). The recollection contrast is defined by

$$c_{\text{recollection}} = {}^{1}\!/_{2}((\beta_{s|s}+\beta_{d|d})-(\beta_{d|s}+\beta_{s|d}))$$

and the familiarity contrast is

$$c_{\text{familiarity}} = {}^{1}\!/_{2}(\beta_{d|s}+\beta_{s|d})-\beta_{n|n},$$

where the β are the estimated regression coefficients and the subscripts refer to the five conditions (s|s, d|s, s|d, d|d, n|n). A second-level mixed-effects analysis was carried out after normalization of images to MNI space. Group statistical maps were computed for the contrasts familiarity and recollection, as defined previously, with a family-wise error rate (FWE) < 0.05. Cluster significance was computed using Monte-Carlo simulations in AFNI (Cox, 1996).

Results

Simulation 1 using block designs

Block designs have been shown to have optimal detection power for simple contrasts of the form A − B, where A and B refer to the amplitudes of the corresponding stimuli sequences. Though block designs are not suitable for the proposed memory task because the arrangement of stimuli can only be balanced up to first order and thus leads to predictability of the stimuli, there is theoretical interest in the detection power of such a block arrangement for more complicated contrasts involving probabilistic responses. We carried out simulations for the memory task by specifying a block-type arrangement of the "same", "different", and "new" stimuli for block sizes of 1 to 30 stimuli, where each stimulus lasts 3 s. The arrangement of stimuli was balanced to first order (i.e. same number of "same" stimuli, "different" stimuli, and "new" stimuli). Using the mean behavioral probabilities from the pilot study (see section Materials and methods or Table 1), 100 different realizations of the five conditions (s|s, d|s, s|d, d|d, n|n) were chosen for each block size and the median detection power for the combined contrast (0.5 ∗ (familiarity + recollection)) was computed (Fig. 2). The best detection power was obtained for a block size of 4 stimuli (12 s).

Simulation 2 using random designs

Designs determined by a random number generator are the easiest to implement and are non-predictable to any order. However, random designs usually suffer from a low contrast detection power (Liu et al., 2001). In our case, it is not clear without simulation to predict the detection power of random designs for the memory task when probabilistic behavioral information is included. Since the contrasts of interest (recollection and familiarity) depend on the behavioral performance (parameterized by random variables) of each subject, random designs are not necessarily ideal for probabilistic tasks similar to ours. We tested 500 random designs at each generation (without invoking any optimization) leading to 50,000 different random

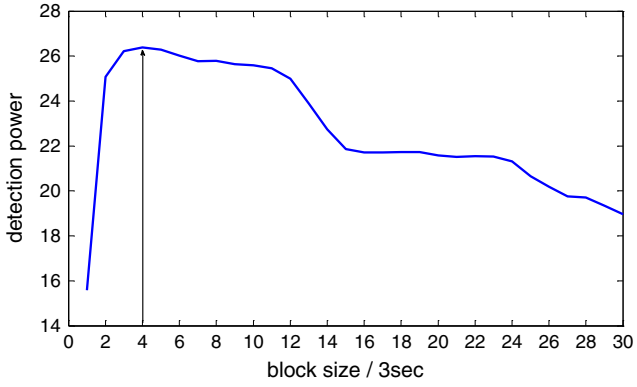

Fig. 2. Median detection power of the proposed memory task assuming a block-type arrangement of the stimuli sequences with equal block size for "same" stimuli, "different" stimuli, and "new" stimuli. The simulation was carried out using the mean behavioral probabilities determined by pilot data (see Table 1, last line). Note, each stimulus has a duration of 3 s and the optimal detection power occurs at a block size of 4 stimuli equaling a duration of 12 s.

configurations. However, the median detection power of the best random configuration did not exceed 24.9, a value which is about 6% lower than the best block-type design (see Fig. 3, top). The random design was not explicitly controlled for non-predictability.

Simulation 3 using optimized designs

Starting with a random design, we determined the optimal design using the proposed genetic algorithm without any non-predictability constraint. The obtained solution vector had non-predictability indices of 0.786, 0.577, and 0.251 for the orders 1, 2, and 3, respectively (Fig. 3). Although, this scenario is less realistic for the given recognition memory paradigm, it provides an upper limit of the contrast detection power which can never be surpassed by any arrangement of the stimuli sequence, and thus has important theoretical value. The genetic algorithm converged at generation 100 with 99% of the maximum achieved (at generation 1000). In comparison to the best random design, the optimal design achieved 28% increased contrast detection power, and in comparison to the best block design, the optimal design achieved 18% increased detection power (Fig. 3 top). For the optimized design, we computed the distribution of the stimuli durations (referred to as block-lengths) and found that the distribution for the "same" and "different" stimuli is pseudo-random (most-likely block length = 3 s) whereas for the "new" stimuli the distribution is block-like (most-likely block length = 9 s) (Fig. 3 middle and bottom).

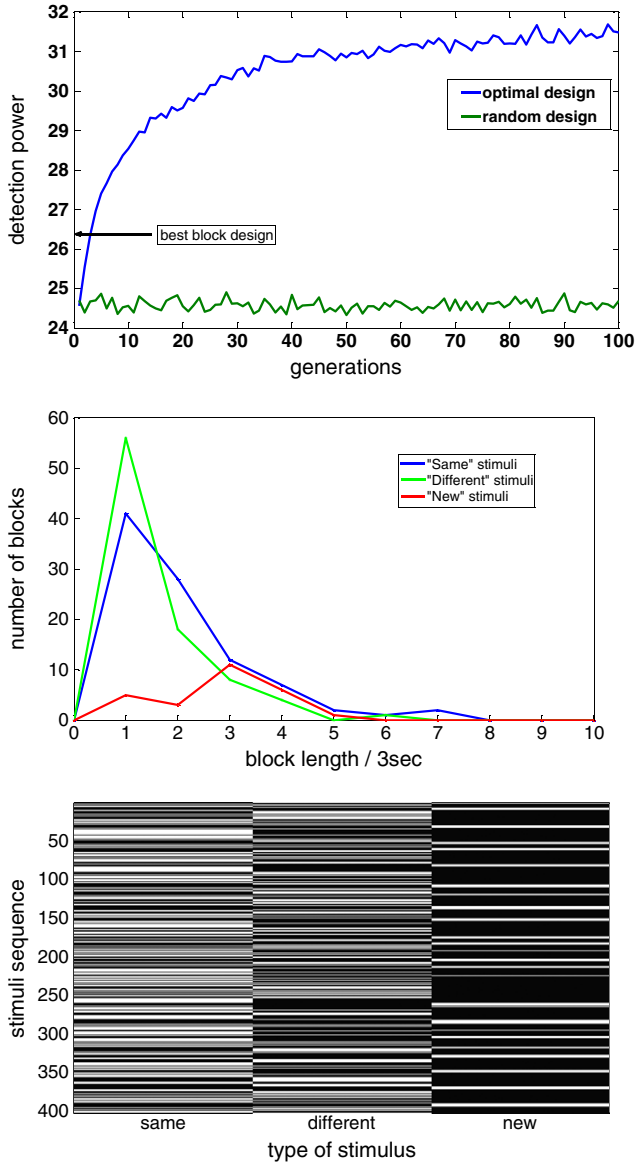

Fig. 3. Top: Optimization of median contrast detection power for the proposed memory task without any non-predictability constraint. The obtained solution vector had non-predictability indices of 0.786, 0.577, and 0.251 for the orders 1, 2, and 3, respectively. The genetic algorithm converged at generation 100 with 99% of the maximum achieved (at generation 1000) (see blue curve). For comparison, the detection power was calculated using 50,000 random arrangements of stimuli (500 random configurations at each generation) (see green curve). Note that the best block design (Fig. 2) is better than the best random design but still about 20% inferior to the optimal design. Middle: Distribution of the block length of the optimal design for the three stimuli. Bottom: Optimal stimulus sequence for "same" stimuli, "different" stimuli, and "new" stimuli. Note that the arrangement of the "new" stimuli are block-like (most-likely block size is 3 stimuli (9 s)), as determined by the genetic algorithm.

Please note that the convergence is not increasing in a monotone way. Due to the probabilistic loop, the detection power is a random variable with a mean and a standard deviation. According to our simulations, this standard deviation of is about 2 for the probabilistic loop, almost independent of the number of generations. At generation 100 using 100 trials for the probabilistic loop, the standard deviation of the median detection power is about 0.12. If the number of trials is increased to say 10,000, the standard deviation of the median detection power at generation 100 is reduced to 0.01 according to our simulations. Thus, there will always be oscillations in the convergence performance because of the inner loop (which is probabilistic). Without the probabilistic loop, the detection power is a monotone increasing function.

We repeated the analysis under the more realistic experimental condition that all stimuli are approximately balanced for the first three orders (non-predictability indices $\geq$ {0.975, 0.9, 0.85}) (Fig. 4). These were the same constraints used to generate the condition sequences that were actually used in the experiment. The genetic algorithm converged at generation 100 with 99% of the maximum achieved (at generation 1000). In comparison to the best random design, the optimal design achieved about 8% increased contrast detection power (26.5 for the optimal design from Fig. 4 top and 24.5 for the random design from Fig. 3 top). For the optimized design, we computed the distribution of the stimuli durations and found that the distribution for all three stimuli is pseudo-random (most-likely block length = 3 s) (Fig. 4 middle and bottom).

We also carried out simulations for intermediate scenarios. If it would be sufficient to have a perceived randomness with non-predictability indices of the first 3 orders $\geq$ {0.975, 0.8, 0.75} or {0.975, 0.6, 0.55}, our simulations indicate an improvement of 9% to 13%, respectively, compared to a random design. The stimuli sequences for both of these scenarios are pseudo-random with similar appearance to the more general case of Fig. 4.

Computation time

Typical computation time for the simulations using MATLAB on a computer equipped with Intel Core 2, 2.4 GHz CPU, and 4 GB memory was about 20 min per generation of the proposed genetic algorithm of Fig. 1. Note that the inner loop of the genetic algorithm was executed 100 times for each generation to compute the median contrast detection power based on the expected behavioral probabilities.

Robustness of design against misspecification

It is important to investigate the robustness of the obtained contrast detection power of our design, which was not optimized *for each individual subject* but optimized based on *the average pilot data*, against misspecification of the probabilities that actually occurred during fMRI scanning (Table 2). We have computed the ratio of the detection power for the misspecified design and the optimal design

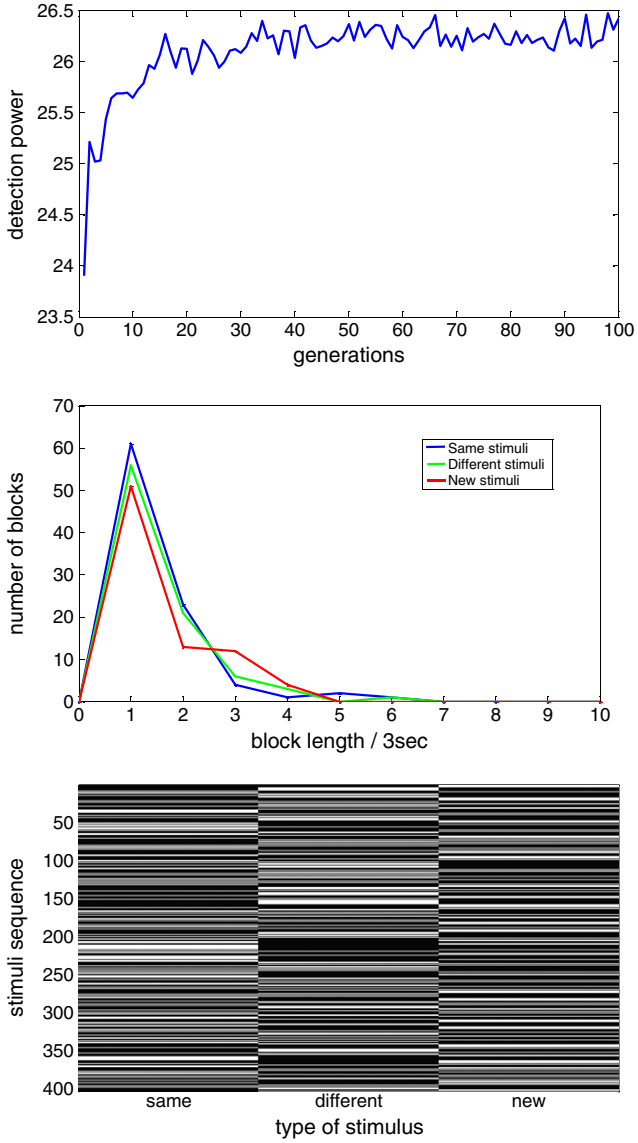

Fig. 4. Top: Optimization of median detection power for the proposed memory task under the experimental condition that stimuli are approximately balanced for 1., 2., and 3. order (non-predictability indices $\geq$ {0.975, 0.9, 0.85}). The obtained solution vector had non-predictability indices of 0.978, 0.908, and 0.853 for the orders 1, 2, and 3, respectively. The genetic algorithm converged at generation 100 with 99% of the maximum achieved (at generation 1000). Middle: Distribution of the block length of the optimal design for the three stimuli. Bottom: Optimal stimulus sequence for "same" stimuli, "different" stimuli, and "new" stimuli.

for each subject data. Furthermore, we list the ratio of the detection power for the random design and the optimal design. The results listed in Table 2 show that our approach based on average pilot data was 97% effective in obtaining the maximum detection power whereas a random approach would have been only 94% effective. These results indicate that on average we were able to achieve a 3% improvement of the contrast detection power (compared to a random design) with behavioral probabilities from average pilot data. This improvement may seem low. Nevertheless, the methods introduced in this article can be applied to *individual* subject pilot data instead of *average* pilot data, which will increase the detection power by 2% to 14% (mean 6%, std 3%) over a random design (computed from Table 2).

Note, that the detection power is not easily predictable based on the behavioral probabilities. For example, subject 14 is farthest from optimal whereas other subjects (e.g., 16) seem to have behavioral

Table 2
Robustness of design against misspecification of behavioral probabilities and comparison to a random design.

Subject #	$\frac{\xi(\text{misspecified})}{\xi(\text{optimal})}$	$\frac{\xi(\text{random})}{\xi(\text{optimal})}$
1	0.977	0.931
2	0.966	0.934
3	0.975	0.908
4	0.967	0.948
5	0.978	0.969
6	0.984	0.950
7	0.968	0.920
8	0.965	0.948
9	0.970	0.968
10	0.979	0.943
11	0.969	0.944
12	0.994	0.972
13	0.975	0.920
14	0.881	0.872
15	0.978	0.958
16	0.991	0.985
17	0.987	0.974
18	0.942	0.942
Mean	0.969	0.944
std	0.025	0.027

Note: The term ξ(misspecified) is the detection power that was achieved during fMRI scanning based on the actual behavioral response of the subject and a stimulus arrangement that was optimized for the mean behavioral probabilities from pilot data. The term ξ(optimal) is the theoretical detection power if the actual behavioral response would have been used to optimize the stimulus sequence. The term ξ(random) is the best detection power of 10,000 random designs.

results that deviate more from the pilot averages. This odd behavior may be explained by the fact that subject 16 has a larger imbalance between s|s + d|d and d|s + s|d responses. In fact, the responses for d|s + s|d are almost twice as large in subject 16 compared to subject 14. This imbalance leads to a double penalty in subject 16 because we optimize for two different contrasts, but both contrasts involve the term d|s + s|d. The result is that the optimal detection power ξ is low for subject 16, and ξ(optimal) and ξ(random) do not differ much.

Further optimization can be achieved by calibrating task difficulty such that the behavioral probabilities become more balanced (or optimal) for the contrasts of interests. We carried out simulations with more balanced behavioral probabilities for recollection and familiarity (Table 3). Note that even a small increase in the difficulty level of the task can lead to a significantly increased detection power. From this perspective, it seems that stimuli should be carefully selected for each contrast of interest. However, we acknowledge that this is not always possible due to conflicting goals and practicality of the experiment. For example, orientation recollection is near chance in the bottom row of Table 3, which would not be optimal from a psychological perspective.

fMRI results

In Fig. 5, we show group activation maps for the familiarity and recollection contrasts at an individual (single voxel, unadjusted)

Table 3
Contrast detection power as a function of the behavioral probabilities.

p(s|s)	p(d|s)	p(s|d)	p(d|d)	p(n|n)	ξ(random)	ξ(optimal)	$\frac{\xi(\text{optimal})}{\xi(\text{random})}$
0.80	0.20	0.30	0.70	0.93	35.7	37.7–39.7	1.056–1.112
0.75	0.25	0.35	0.65	0.93	40.8	43.4–45.9	1.064–1.125
0.70	0.30	0.40	0.60	0.93	45.1	47.9–51.0	1.062–1.131
0.65	0.35	0.45	0.55	0.93	48.3	52.1–55.6	1.079–1.151
0.60	0.40	0.50	0.50	0.93	50.8	55.1–58.5	1.085–1.152

Note: The simulations of ξ(optimized) were performed using the proposed genetic algorithm with non-predictability indices for 1. to 3. order $\geq$ {0.975, 0.9, 0.85}) (lower number in columns) and also for {0.975, 0.6, 0.55}) (higher number in columns). The random design was not controlled for their non-predictability index.

p-value = 0.001. All clusters with a cluster size of at least 424 mm^3 are significant at FWE<0.05, as determined by AlphaSim in AFNI (Cox, 1996). Strong activations are present for recollection > 0 in the anterior medial prefrontal cortex, the lateral parietal cortex, the lateral temporal cortex, hippocampus, parahippocampal gyrus, and posterior cingulate cortex. For familiarity > 0, the strongest activations occur in the lateral anterior prefrontal cortex, the dorsolateral prefrontal cortex, the superior parietal cortex, the caudate nucleus, precentral gyrus, and peristriate (mid occipital) area. Most activations are bilateral with larger cluster sizes in the left hemisphere.

All significantly activated regions at FWE<0.05 are listed in Table 4 (recollection) and Table 5 (familiarity). The entries in Tables 4 and 5 indicate that there are many overlapping areas between recollection and familiarity when both positive and negative contrasts are taken into account. The reason for overlapping areas is likely related to the common $\beta_{d|s} + \beta_{s|d}$ term in the recollection and familiarity contrast. Regions that show both positive recollection activations and negative familiarity activations might reflect confidence differences between conditions insofar as both $\beta_{s|s} + \beta_{d|d}$ and $\beta_{n|n}$ should be associated with higher confidence. From this perspective, focusing on activations that are unique to either familiarity or recollection irrespective of the sign might better reflect true differences between familiarity and recollection. Major clusters (positive or negative) of the familiarity contrast (but not for the recollection contrast) occur in the right prefrontal cortex (Frontal Inf Oper, Frontal Inf Tri, Frontal Mid Tri, Frontal Sup), right caudate, bilateral Heschl gyrus, bilateral mid occipital cortex, left superior and inferior occipital cortex, bilateral inferior parietal cortex, left superior parietal cortex, left precentral cortex, right inferior temporal cortex, left superior temporal pole, and bilateral mid temporal cortex. For the recollection contrast (but not the familiarity contrast), major activations are in the left cerebellum, left lingual gyrus, left pallidum, right putamen, and right thalamus.

Discussion

Design optimization

The purpose of this study was to determine if the stimuli sequence for a recognition task can be optimized for maximum detection power using probabilistic behavioral information. Our analysis focused on implementing anticipated behavioral probabilities in a genetic algorithm to achieve the best design for contrast detection power of familiarity and recollection. We clearly showed that behavioral probabilities obtained from independent pilot data are helpful for design optimization.

We also examined stimuli sequences based on block-type designs and found that block-type designs are not necessarily optimal, even if non-predictability criteria of the behavioral outcomes are not implemented. As in our case, where the contrast is more complicated, a block design will not be the best design for the arrangement of stimuli. Nevertheless a particular stimulus-type can have a block-like appearance, as shown for the “new” stimuli but not for the “same” and “different” stimuli. We attribute this finding to the particular contrasts of interest (0.5 ∗ (familiarity + recollection)) and the high probability for the correctly identified “new” stimuli.

We explored the usefulness of a random design without invoking the genetic algorithm and found that a random design is suboptimal for paradigms similar to our recognition task. Despite the probabilistic nature of the regressors, a genetic algorithm can improve the design by significantly increasing detection power. Results will be more accurate when the anticipated behavioral responses are closer

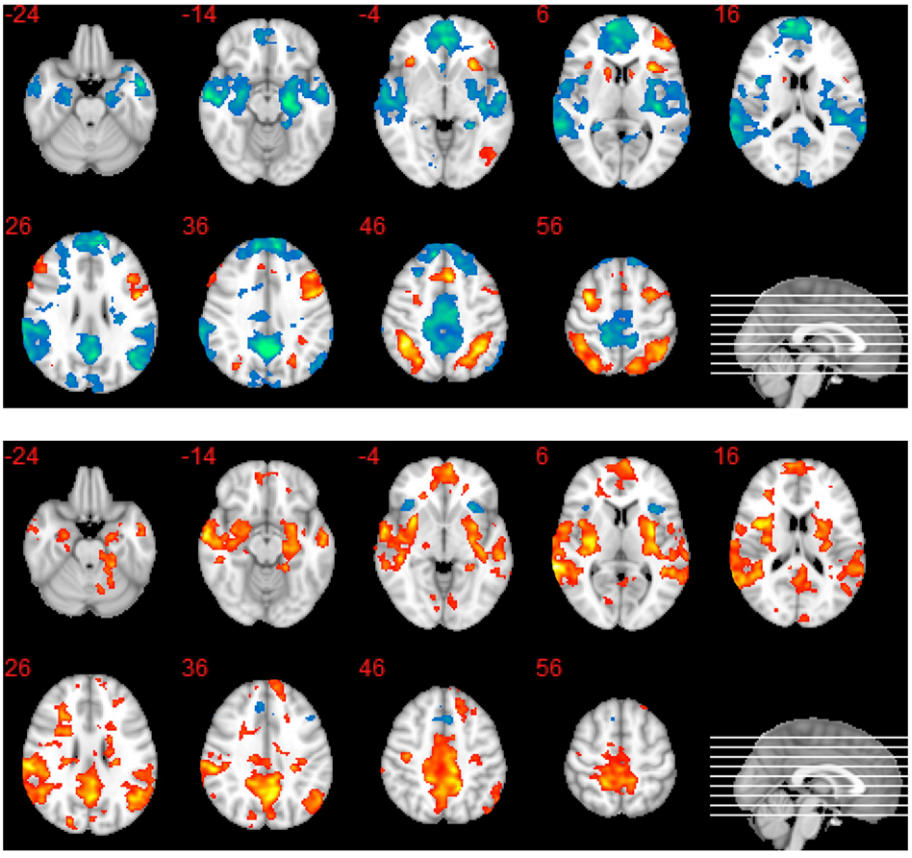

Fig. 5. Group activation maps for familiarity > 0 (top) and recollection > 0 (bottom) at an individual (single voxel, unadjusted) p-value = 0.001. All clusters with a cluster size of at least 424 mm^3 are significant at FWE<0.05. Images are in radiological convention (left is right and vice versa).

Table 4
Significant regions (corrected $p<0.05$) for recollection contrast.

MNI region	Cluster size	t-values at peak	MNI coordinates at peak
Amygdala (L;R)	130;167	5.3; 5.2	−28,2,−12; 32,−2,−10
Angular (L;R)	630;401	5.9;5.3	−52,−60,28;48,−54,26
Calcarine (L;R)	214;65	4.5;3.8	−2,−64,22;2,−58,18
Caudata (L)	60	4.7	−20,0,24
Cerebellum_4_5 (L)	165	5.3	−18,−42,−26
Cerebellum_6 (L)	146	4.2	−22,−54,−24
Cingulum_Ant (L)	144	4.6	−4,54,0
Cingulum_Mid (L;R)	1139;1047	6.7;5.5	−12,−40,40;12,−46,36
Cingulum_Mid (R)	83	−4.4	4,34,38
Cingulum_Post (L;R)	373;198	6.2;5.4	−8,−52,34;4,−54,32
Cuneus (L;R)	325;161	5.2;5.0	0,−70,32;18,−84,26
Frontal_Inf_Tri (L)	80	−4.3	−34,20,12
Frontal_Mid (L)	256	5.6	−22,32,44
Frontal_Med_Orb (L;R)	350;329	5.4;5.0	−8,56,−6;8,44,−6
Frontal_Sup (L)	471	5.4	−20,32,44
Frontal_Sup_Medial (L;R)	697;379	5.3;5.2	−6,66,8;8,64,14
Frontal_Sup_Medial (L)	89	−5.5	−6,24,42
Fusiform_(L)	143	5.7	−22,−42,−12
Hippocampus (L;R)	445;383	5.0;4.9	−28,−22,−14;34,−16,−14
Insula (L;R)	331;442	5.9;6.1	−32,2,10;38,2,16
Insula (L;R)	330;174	−5.6;−4.8	−30,20,−2;32,24,−4
Lingual (L;R)	174;131	5.3;4.2	−24,−42,−8;14,−78,2
Occipital_Sup_(R)	129	5.1	20,−84,26
Pallidum_(L)	68	4.4	−22,4,−2
Paracentral_Lobule (L;R)	205;151	5.5;5.4	−6,−28,52;16,−40,50
ParaHippocampal (L;R)	255;168	4.7;4.5	−24,−40,−8;24,−8,−24
Parietal_Sup_(R)	177	5.4	16,−46,56
Postcentral (L;R)	190;359	4.5;5.2	−38,−24,38;16,−38,70
Precentral_(R)	68	5.5	38,−18,42
Precuneus (L;R)	1614;1182	7.2;5.8	−6,−58,32;2,−54,34
Putamen (L;R)	656;611	6.3;6.8	−32,−2,6;28,−12,10
Rolandic_Oper (L;R)	107;488	4.0;5.9	−54,0,10;42,4,14
Supp_Motor_Area (L;R)	155;278	4.4;4.7	−6,−16,50;10,−18,50
Supp_Motor_Area (L)	63	−4.8	−6,20,46
SupraMarginal (L;R)	290;1001	5.1;6.5	−58,−40,30;54,−30,32
Temporal_Inf_(L)	71	4.4	−56,−4,−28
Temporal_Mid (L;R)	1769;2089	7.0;7.8	−46,−44,10;66,−40,6
Temporal_Pole_Sup_(R)	61	4.8	62,4,−8
Temporal_Sup (L;R)	855;1832	6.7;4.8	−60,−42,14;56,−40,12
Thalamus_(R)	93	5.5	22,−22,10

Note: Cluster volume is given by cluster size*8 mm^3

Table 5
Significant regions (corrected $p<0.05$) for familiarity contrast.

MNI region	Cluster size	t-values at peak	MNI coordinates at peak
Amygdala (L;R)	158;181	−5.7;−6.0	−28,2,−16;24,−6,−14
Angular_(R)	318	5.5	40,−56,48
Angular (L;R)	391;485	−7.3;−6.4	−44,−60,28;60,−60,28
Calcarine (L)	223	−4.9	−2,−64,22
Caudate (L;R)	56;101	4.2;4.6	−10,14,4;12,14,8
Cerebellum_Crus1_(R)	88	4.7	40,−66,−30
Cingulum_Ant (L;R)	527;436	−7.5;−8.7	0,44,2;2,46,4
Cingulum_Mid (L;R)	1078;1026	−7.9;−8.9	−12,−52,34;4,−48,38
Cingulum_Post (L;R)	277;140	−10.5;−8.7	−4,−52,34;4,−54,32
Cuneus (L;R)	547;234	−6.3;−4.8	−8,−60,26;18,−84,28
Frontal_Inf_Oper (L;R)	221;151	5.8;4.3	−46,22,34;62,14,22
Frontal_Inf_Orb (L)	54	−4.5	−36,18,−18
Frontal_Inf_Tri (L;R)	500;185	5.7;4.6	−46,20,30;52,38,28
Frontal_Inf_Tri (R)	66	−4.0	54,34,6
Frontal_Mid (L;R)	461;353	5.7;7.2	−32,4,60;30,−2,54
Frontal_Mid (L;R)	600;428	−5.9;−4.9	−24,30,44;22,46,30
Frontal_Mid_Orb (L)	65	−4.2	−26,38,−12
Frontal_Med_Orb (L;R) (L;R)	467;537	−7.9;−7.8	2,58,−8;2,58,−6
Frontal_Sup (L;R)	64;254	4.9;5.9	−26,−2,62;28,10,62
Frontal_Sup (L;R)	1001;725	−8.1;−8.0	−14,60,20;16,56,14
Frontal_Sup_Medial (L)	183	5.5	−6,22,44
Frontal_Sup_Medial (L;R)	1622;1244	−8.5;−10.0	−12,60,20;14,58,12
Fusiform (L)	135	−6.8	−22,−42,−12
Heschl (L;R)	135;53	−7.8;−4.8	−38,−22,6;58,−2,8
Hippocampus (L;R)	499;403	−9.4;−7.4	−24,−16,−16;24,−12,−18
Insula (L;R)	359;119	7.3;4.8	−32,18,4;34,26,−4
Insula (L;R)	525;602	−6.7;−7.3	−38,−18,2;38,6,10
Lingual (R)	103	−4.7	14,−80,0
Occipital_Mid (L;R)	115;92	4.8;4.3	−28,−64,40;30,−70,36
Occipital_Inf_(L)	185	4.0	−52,−72,−4
Occipital_Sup_(R)	184	6.1	24,−66,48
Occipital_Sup (L;R)	137;157	−4.9;−4.8	−12,−92,34;18,−82,30
Paracentral_Lobule (L;R)	147;167	−5.3;−6.7	0,−30,54;10,−30,50
ParaHippocampal (L;R)	371;336	−7.1;−8.0	−26,−40,−8;22,−14,−20
Parietal_Inf (L;R)	1061;621	7.5;8.5	−40,−54,50;40,−50,50
Parietal_Sup (L;R)	918;907	6.2;6.2	−20,−66,50;24,−66,50
Postcentral (L;R)	118;159	−4.3;−6.0	−24,−40,62;16,−38,70
Precentral (L;R)	862;177	7.4;7.3	−38,2,62;30,−2,52
Precuneus (L;R)	158;118	5.1;4.5	−14,−70,58;12,−66,64
Precuneus (L;R)	1285;998	−11.0;−10.6	−8,−54,32;2,−54,34
Putamen (L)	387	−5.3	−30,2,10
Rectus (R)	62	−4.9	8,50,−14
Rolandic_Oper (L;R)	229;390	−5.3;−5.6	−40,−34,16;54,−12,22
Supp_Motor_Area (L;R)	241;88	6.2;4.3	10,22,52;−6,22,44
Supp_Motor_Area (L;R)	143;176	−5.4;−5.2	−6,−16,50;10,−26,52
SupraMarginal (L;R)	285;891	−5.5;−8.0	−62,−34,30;66,−42,28
Temporal_Inf (L;R)	326;287	−5.5;−6.3	−48,6,−34;54,0,−36
Temporal_Mid (L;R)	1561;1999	−9.2;−9.0	−54,2,−24;60,−10,−14
Temporal_Pole_Mid (L;R)	89;145	−5.3;−4.8	−50,10,−30;52,8,−32
Temporal_Pole_Sup (L;R)	130;110	−6.1;−4.6	−36,16,−20;44,4,−16
Temporal_Sup (L;R)	1606;2161	−7.7;−8.0	−40,−20,2;62,−38,12

Note: Cluster volume is given by cluster size*8 mm^3

to the actual responses during fMRI scanning than to previous measurements. The detection power can be increased even further if the number of responses for the different conditions becomes more similar, which could be achieved by manipulating the difficulty of the task.

The trade-off in obtaining the best design with the largest detection power is the perceived randomness. For our experimental setup, we used a design that was balanced for orders 1 to 3, which reduced the theoretically achieved detection power. If it would be sufficient to reduce the perceived randomness to {0.975, 0.8, 0.75} or {0.975, 0.6, 0.55} for the first 3 orders, our simulations indicate an improvement of 9% to 13% compared to a random design.

For the simple case of using behavioral probabilities from average pilot data, we showed that our design with the proposed genetic algorithm is robust against design misspecification, which arises due to the fact that actual subject performance during fMRI scanning may be different than the pilot data. We obtained for all subjects an increase of the detection power compared to a random design, though we did not optimize the experimental design for each individual subject but used average pilot data. This result can be attributed to the closeness of the behavioral probabilities at scanning to the pilot data.

Effect of imbalance of task responses

We obtained a larger imbalance of the responses for conditions involving the familiarity contrast (#(d|s)+#(s|d) vs. #(n|n)) than for recollection contrast (#(s|s)+#(d|d) vs. #(d|s)+#(s|d)) leading to a small bias in the ability to detect activations for familiarity vs. recollection. Using a 2-sample *t*-test, the significance (*t*-value) is approximately proportional to

$$s = \frac{1}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}},$$

where n_1 and n_2 are the number of responses for the contrasting conditions ($n_{\downarrow}1$=#(d|s)+#(s|d) and for familiarity; $n_{\downarrow}1$=#(s|s)+#(d|d) and $n_{\downarrow}2$=#(d|s)+#(s|d) for recollection), assuming similar variance of the samples. Using Table 1, we computed *s* for familiarity and recollection of all subjects, and obtained, on average, a 6% increased *s* for recollection over familiarity, with a standard deviation of 2%. It follows that the given imbalance of the behavioral probabilities leads

to more significant activation of the recollection contrast than the familiarity contrast.

Comparison with other fMRI studies

A more detailed discussion of the present whole brain analysis of familiarity and recollection can be found elsewhere (Herzmann et al., submitted for publication). There, we report a comprehensive analysis of the present fMRI activation, and a parallel event-related potential study which we conducted within the same subjects. Here, we want to briefly discuss the present results with regard to similar, previous investigations to provide evidence for the successful use of the proposed optimization algorithm.

Previous whole-brain studies on familiarity and recollection measured by item and source memory judgments, respectively, reported very similar results as found in the present experiment. The present finding of stronger activation for the familiarity contrast in the prefrontal, occipital, and parietal cortex corresponds well with previous findings (e.g., Cansino et al., 2002; Dobbins et al., 2003; Ragland et al., 2006; Skinner and Fernandes, 2007; Slotnick et al., 2003; Wheeler and Buckner, 2004). The present recollection activations are in line with previous studies that reported activation in the thalamus, amygdala, and the lingual cortex for the recollection contrast (reviewed in Spaniol et al., 2009).

Similar results as in the present study were also found in previous investigations that measured recollection and familiarity with subjective memory judgments. Henson et al. (1999) used the remember–know procedure, originally introduced by Tulving (1985), to study recollection and familiarity of words. Activations for familiarity >0 and recollection <0 were found at right (lateral and medial) prefrontal cortex. This result is in strong agreement with the current study where several regions of the right prefrontal cortex were activated. Activity of the right prefrontal cortex can be explained by a larger working memory demand when memory judgments are only familiar and thus appear less certain and more difficult for the participant. This observation has been called "adoption of retrieval mode" for familiarity decisions (Kapur et al., 1995; Nyberg et al., 1995). A consequence is that the response times for familiarity decisions are increased compared to recollection decisions.

Activations for recollection contrast >0 reported by Henson et al. (1999) were in left parietal, left prefrontal, and posterior cingulate. Except for the left parietal activation, these findings agree well with our data. When Spaniol et al. (2009) considered differences between "subjective recollection" (e.g., remember/know judgments) and objective recollection (e.g., source memory judgments or orientation recollection), left inferior parietal activation was more strongly associated with subjective than objective recollection. For familiarity contrast <0, Henson et al. (1999) found bilateral amygdala and bilateral temporo-occipital regions active just as in our study. Significance of this negative contrast is expected presumably due to a response to novel stimuli involving the anterior portion of the medial temporal lobes (amygdala and hippocampus).

The study by Yonelinas et al. (2005) used a modified remember–know procedure where subjects rated the confidence of their familiarity judgments for words that they knew they had studied before but which they did not remember. Recollection-related regions were identified for the contrast remember $>$ most confident familiarity ratings and found to be in bilateral hippocampus, anterior medial prefrontal cortex, lateral parietal and temporal cortex, posterior cingulate cortex, and left parahippocampal cortex. Familiarity-related regions were determined as those correlated with increasing familiarity confidence ratings from least to most confident and found to be located in anterior lateral prefrontal cortex, dorsolateral prefrontal cortex, superior lateral parietal cortex, and precuneus. We obtained very similar activations and our results agree in general well with the study by Yonelinas et al. A few differences, however, exist. One difference is that in our study precuneus is activated for both recollection and familiarity (and not just for familiarity). Also, we found amygdala to be active for recollection >0 but not for familiarity >0. Furthermore, the caudate nucleus is clearly activated in our study for familiarity but not for recollection. Since the level of significance is lower for the familiarity >0 contrast, we also investigated the familiarity activation map at a more liberal t-threshold (uncorrected $p=0.01$) to see if the pattern of positive activation changes significantly. We did not find any evidence that more overlap occurs between the two contrasts at the lower threshold. In particular, we did not find any activation in the medial temporal lobes (hippocampus, parahippocampal gyrus, entorhinal cortex, perirhinal cortex, amygdala) for familiarity >0. This important result also agrees very well with Yonelinas et al. (2005), except that our study cannot address memory strength (or confidence) as a possible confound (also see Wixted and Squire, 2011). Furthermore, we did not find any indication that the perirhinal cortex is involved in familiarity, contrary to current opinion (Diana et al., 2007; Haskins et al., 2008). A reason for this discrepancy may be signal drop-out due to susceptibility effects of the sphenoid sinus affecting the anterior part of the parahippocampal gyrus (entorhinal cortex, perirhinal cortex). Signal drop-out in these regions is especially strong for axial acquisitions of echoplanar data (Jin et al., in press).

Conclusions

In this article, we propose a genetic algorithm that includes probabilistic behavioral information to optimize the design of a task for contrast detection power. We have applied this optimization technique to a recognition memory task to investigate familiarity and recollection of pictures of common objects with different orientations. We have shown that the order of stimuli can be optimized for probabilistic behavioral responses, leading to better contrast detection power than a random design or the best block design. Furthermore, the optimized design is robust to small changes of the behavioral probabilities, which occur during actual fMRI scanning due to differences in the subjects' performance from the pilot data. Contrast detection power can be further increased by optimizing the task design for each individual subject. The present genetic algorithm can be applied to any case in which fMRI contrasts are dependent on probabilistic responses that can be estimated from pilot data.

Acknowledgments

This research was supported by the NIH/NIA (grant number 1R21AG026635), NIH/NIMH (grant number R01MH64812), and a University of Colorado Council for Research & Creative Work (CRCW) Faculty Sabbatical Fellowship to Tim Curran.

References

Aggleton, J.P., Brown, M.W., 1999. Episodic memory, amnesia, and the hippocampal–anterior thalamic axis. Behav. Brain Sci. 22, 425–489.
Ahn, C.W., 2006. Advances in evolutionary algorithms: theory, design and practice. Studies in Computational Intelligence. Springer, Berlin, New York.
Birn, R.M., Cox, R.W., Bandettini, P.A., 2002. Detection versus estimation in event-related fMRI: choosing the optimal stimulus timing. NeuroImage 15, 252–264.
Buckner, R.L., 1998. Event-related fMRI and the hemodynamic response. Hum. Brain Mapp. 6 (5–6), 373–377.
Cansino, S., Maquet, P., Dolan, R.J., Rugg, M., 2002. Brain activity underlying encoding and retrieval of source memory. Cereb. Cortex 12, 1048–1056.
Carr, V.A., Rissman, J., Wagner, A.D., 2010. Imaging the human medial temporal lobe with high-resolution fMRI. Neuron 65 (3), 298–308.
Cordes, D., Nandy, R., 2007. Independent component analysis in the presence of noise in fMRI. Magn. Reson. Imaging 25, 1237–1248.
Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.
Dale, A.M., 1999. Optimal experimental design for event-related fMRI. Hum. Brain Mapp. 8, 109–114.
Diana, R.A., Yonelinas, A.P., Ranganath, C., 2007. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends Cogn. Sci. 11 (9), 379–386.

Dobbins, I.G., Rice, H.J., Wagner, A.D., Schacter, D.L., 2003. Memory orientation and success: separable neurocognitive components underlying episodic recognition. Neuropsychologia 41, 318–333.
Eichenbaum, H., Yonelinas, A.P., Ranganath, C., 2007. The medial temporal lobe and recognition memory. Annu. Rev. Neurosci. 30, 123–152.
Frackowiak, R.S.J, editor-in-chief, 2004. Human Brain Function, 2nd edition, Elsevier Science, San Diego.
Friston, K.J., Zarahn, E., Josephs, O., Henson, R.N.A., Dale, A.M., 1999. Stochastic design in event-related fMRI. NeuroImage 10, 607–619.
Friston, K.J., Josephs, O., Zarahn, E., Holmes, A.P., Rouquette, S., Poline, J.-B., 2000. To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. NeuroImage 12 (2), 196–208.
Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9, 416–429.
Gonzalez, R.C., Woods, R.E., 1993. Digital Image Processing. Addison-Wesley Publishing Company.
Haskins, A.L., Yonelinas, A.P., Quamme, J.R., Ranganath, C., 2008. Perirhinal cortex supports encoding and familiarity-based recognition of novel associations. Neuron 59, 554–560.
Henson, R.N.A., Rugg, M.D., Shallice, T., Josephs, O., Dolan, R.J., 1999. Recollection and familiarity in recognition memory: an event-related functional magnetic resonance imaging study. J. Neurosci. 19, 3962–3972.
Herzmann, G., Jin, M., Cordes, D., Curran, T., submitted for publication. A within-subject ERP and fMRI investigation of orientation-specific recognition memory for pictures. Cogn. Neurosci.
Jin, M., Pelak, V.S., Cordes, D., in press. Aberrant default mode network in subjects with amnestic mild cognitive impairment using resting-state functional MRI. Magn. Reson. Imaging.
Josephs, O., Henson, R.N., 1999. Event-related functional magnetic resonance imaging: modelling, inference and optimization. Philos. Trans. R. Soc. Lond. B Biol. Sci. 354 (1387), 1215–1228.
Kao, M.-H., Mandal, A., Lazar, N., Stufken, J., 2009. Multi-objective optimal designs for event-related fMRI studies. NeuroImage 44, 849–856.
Kapur, S., Craik, F.I., Jones, C., Brown, G.M., Houle, S., Tulving, E., 1995. Functional role of the prefrontal cortex in retrieval of memories: a PET study. Neuroreport 6, 1880–1884.
Liu, T.T., 2004. Efficiency, power, and entropy in event-related fMRI with multiple trial types. Part II: design of experiments. NeuroImage 21, 401–413.
Liu, T.T., Frank, L.R., 2004. Efficiency, power, and entropy in event-related fMRI with multiple trial types. Part I: theory. NeuroImage 21, 387–400.
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.
Mechelli, A., Price, C.J., Henson, R.N.A., Friston, K.J., 2003. Estimating efficiency a priori: a comparison of blocked and randomized designs. NeuroImage 18, 798–805.
Norman, K.A., O'Reilly, R.C., 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary learning systems approach. Psychol. Rev. 110, 611–646.
Nyberg, L., Tulving, E., Habib, R., Nilsson, L.G., Kapur, S., Cabeza, R., McIntosh, A.R., 1995. Functional brain maps of retrieval mode and recovery of episodic information. Neuroreport 7, 249–252.
Parks, C.M., 2007. The role of noncriterial recollection in estimating recollection and familiarity. J. Mem. Lang. 57, 81–100.
Ragland, J.D., Valdez, J.N., Loughead, J., Gur, R.C., Gur, R.E., 2006. Functional magnetic resonance imaging of internal source monitoring in schizophrenia: recognition with and without recollection. Schizophr. Res. 87, 160–171.
Rugg, M.D., Curran, T., 2007. Event-related potentials and recognition memory. Trends Cogn. Sci. 11, 251–257.
Skinner, E.I., Fernandes, M.A., 2007. Neural correlates of recollection and familiarity: a review of neuroimaging and patient data. Neuropsychologia 45, 2163–2179.
Slotnick, S.D., Moo, L.R., Segal, J.B., Hart Jr., J., 2003. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. Cogn. Brain Res. 17, 75–82.
Spaniol, J., Davidson, P.S.R., Kim, A.S.N., Han, H., Moscovitch, M., Grady, C.L., 2009. Event-related fMRI studies of episodic encoding and retrieval: meta-analyses using activation likelihood estimation. Neuropsychologia 47, 1765–1779.
Squire, L.R., Clark, R.E., Bayley, P.J., 2004. Medial temporal lobe function and memory, In: Gazzinaga, M. (Ed.), The Cognitive Neurosciences, 3rd ed. MIT Press, Cambridge, MA, pp. 691–708.
Squire, L.R., Wixted, J.T., Clark, R.E., 2007. Recognition memory and the medial temporal lobe: a new perspective. Nat. Rev. Neurosci. 8 (11), 872–883.
Tulving, E., 1985. Memory and consciousness. Can. Psychol. 26, 1–12.
Wager, T.D., Nichols, T.E., 2003. Optimization of experimental design in fMRI: a general framework using a genetic algorithm. NeuroImage 18, 293–309.
Wager, T.D., Vazquez, A., Hernandez, L., Noll, D.C., 2003. Accounting for nonlinear BOLD effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies. NeuroImage 18.
Wheeler, M.E., Buckner, R.L., 2004. Functional-anatomic correlates of remembering and knowing. NeuroImage 21, 1337–1349.
Wixted, J.T., Squire, L.R., 2011. The medial temporal lobe and the attributes of memory. Trends Cogn. Sci. 15 (5), 210–217.
Yonelinas, A.P., 2002. The nature of recollection and familiarity: a review of 30 years of research. J. Mem. Lang. 46, 441–517.
Yonelinas, A.P., Jacoby, L.L., 1996. Noncriterial recollection: familiarity as automatic, irrelevant recollection. Conscious. Cogn. 5, 131–141.
Yonelinas, A.P., Otten, J.O., Shaw, K.N., Rugg, M.D., 2005. Separating the brain regions involved in recollection and familiarity in recognition memory. J. Neurosci. 25 (11), 3002–3008.
Yonelinas, A.P., Aly, M., Wang, W.C., Koen, J.D., 2010. Recollection and familiarity: examining controversial assumptions and new directions. Hippocampus 20, 1178–1194.