

# Multisubject fMRI Studies and Conjunction Analyses

K. J. Friston,\* A. P. Holmes,\* C. J. Price,\* C. Büchel,\* and K. J. Worsley†

\*The Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London, WC1N 3BG, United Kingdom; and †Department of Mathematics and Statistics, McGill University, Montreal, Quebec H3A 2K6, Canada

Received February 5, 1999

**In this paper we present an approach to making inferences about generic activations in groups of subjects using fMRI. In particular we suggest that activations common to all subjects reflect aspects of functional anatomy that may be “typical” of the population from which that group was sampled. These commonalities can be identified by a conjunction analysis of the activation effects in which the contrasts, testing for an activation, are specified separately for each subject. A conjunction is the joint refutation of multiple null hypotheses, in this instance, of no activation in any subject. The motivation behind this use of conjunctions is that fixed-effect analyses are generally more “sensitive” than equivalent random-effect analyses. This is because fixed-effect analyses can harness the large degrees of freedom and small scan-to-scan variability (relative to the variability in responses from subject to subject) when assessing the significance of an estimated response. The price one pays for the apparent sensitivity of fixed-effect analyses is that the ensuing inferences pertain to, and only to, the subjects studied. However, a conjunction analysis, using a fixed-effect model, allows one to infer: (i) that every subject studied activated *and* (ii) that at least a certain proportion of the population would have shown this effect. The second inference depends upon a meta-analytic formulation in terms of a confidence region for this proportion. This approach retains the sensitivity of fixed-effect analyses when the inference that only a substantial proportion of the population activates is sufficient.**

© 1999 Academic Press

**Key Words:** functional neuroimaging; fMRI; fixed and random effects; conjunctions; statistical inference

## INTRODUCTION

This paper concerns the detection of common activation foci in multisubject or single-subject multisession functional magnetic resonance imaging (fMRI) studies and addresses the issue of how to make inferences about activation effects when the degree of activation may differ from subject to subject or from session to session. This differential responsiveness, or differential

sensitivity to evoked hemodynamic responses, means that the “fixed-effect” statistical analyses, commonly used in fMRI time-series analysis, may not be appropriate when trying to generalize inferences to the population from which the group was sampled. In what follows we describe one approach to this problem that ensures the validity of inferences while retaining the sensitivity of fixed-effect analyses. This approach uses conjunction analyses (Price and Friston, 1997) and depends upon recent advances in Gaussian field theory (Worsley, 1999).

This paper is divided into two sections. In the first we introduce the distinction between fixed- and random-effect models, in relation to multisubject fMRI studies and differential subject- or session-specific responses. After considering the sources of these “session effects” we review conjunction analyses and how they can be used to make inferences about the particular subject studied and, using a meta-analytic strategy, the population from which they came. The second section is an empirical example of the approach as applied to a fMRI study of evoked responses in the visual system. These data have been used previously to demonstrate the feasibility of multisubject analyses with statistical parametric mapping (Büchel *et al.*, 1997).

## RANDOM- AND FIXED-EFFECT ANALYSES

In Friston *et al.* (1995a) we presented a general framework for the analysis of functional neuroimaging time series that employs the general linear model, to create statistical parametric maps or SPMs and, in Worsley *et al.* (1996), a unified theory of Gaussian fields used in their interpretation. In fMRI data analysis the general linear model is variously known as Anova, AnCova, or multiple linear regression (Friston *et al.*, 1995b). The general linear model has been elaborated to deal with temporal or serial correlations in fMRI time series (Friston *et al.*, 1995b; Worsley and Friston, 1995) and has been applied in the analysis of blocked designs and event-related fMRI (Josephs *et al.*, 1997; Friston *et al.*, 1998). These approaches model the observed hemodynamic responses in terms of a linear

combination of expected effects (i.e., explanatory variables, regressors, or covariates that comprise the “design matrix”) up to the level of autocorrelated residual errors. These effects can range from stimulus waveforms (inferred on the basis of changing task conditions) convolved with a hemodynamic response function to basis functions of peristimulus time in event-related fMRI.

To date the application of the general linear model to fMRI time series has been largely restricted to fixed-effect analyses. These analyses assume that each subject makes the same, fixed contribution to the observed activation and therefore discount random variations from subject to subject. The activation effect is assessed by comparing the contribution of the explanatory variables, in terms of a linear compound or *contrast* of the associated parameter estimates (i.e., regression coefficients), and the residual or error variance to produce a suitable statistic, usually the  $T$  or  $F$  statistic. The problem with using fixed-effect analyses in multisubject designs is acutely apparent when comparing groups of subjects. However, this paper is concerned with the simpler problem of making inferences about activations in a single group of subjects.

For a given group of subjects, there is a fundamental distinction between saying that the average response is significant in relation to the variability *of* the subjects' responses and that there is a significant response in relation to the variability *about* those subject-specific responses. This distinction relates directly to the difference between fixed- and random-effect analyses. The following example tries to make this clear: Consider what would happen if we scanned six subjects with fMRI during the performance of a single task, relative to a baseline. We then constructed a statistical model, in which task-specific effects were modeled separately for each subject. Unknown to us, only one of the subjects activated a particular brain region. When we examine the contrast of parameter estimates, assessing the mean activation over all the subjects, we see that it is greater than zero by virtue of this subject's activation. Furthermore because that model fits the data extremely well (modeling no activation in five subjects and a substantial activation in the sixth) the error variance, on a scan-to-scan basis, is small and the  $T$  statistic is very significant. Can we then say that the group shows an activation? On the one hand we can say, quite properly, that the mean group response embodies an activation but clearly this does not constitute an inference that the group's response is significant (i.e., that this sample of subjects shows a consistent activation). The problem here is that we are using the *scan-to-scan* error variance and this is not necessarily appropriate for an inference about group responses. In order to make the inference that the group showed a significant activation one would have to assess the

variability in activation effects from *subject to subject* (using the contrast of parameter estimates for each subject). This variability now constitutes the proper error variance. In this instance the variance of these six measurements would be large relative to their mean and the corresponding  $T$  statistic would not be significant.

The distinction between the two approaches above relates to how one computes the appropriate error variance. The first represents a fixed-effect analysis and the second a random-effect analysis. In the former the error variance is estimated on a scan-to-scan basis, assuming that each scan represents an independent observation (ignoring serial correlations). Here the degrees of freedom are essentially the number of scans (minus the rank of the design matrix). Conversely, in random-effect analyses, the appropriate error variance is based on the activation from subject to subject in which the effect per se constitutes an independent observation and the degrees of freedom fall dramatically to the number of subjects (minus one). The term random-effect indicates that we have accommodated the randomness of differential responses by comparing the mean activation to the variability in activations from subject to subject. Both analyses are perfectly valid but only in relation to the inferences that are being made.

Let us assume that we want to make inferences that generalize to the population from which the subjects studied were taken. The problem then is to choose between a fixed-effect and a random-effect analysis. On the one hand fixed-effect analyses will be more sensitive because they have substantially more degrees of freedom and the scan-to-scan variability is typically much less than the subject-to-subject response variability. On the other hand, if there are random effects, the fixed-effect analysis will be inappropriate and is likely to be capricious if we want to generalize the inference to the population from which our cases came. We have already described the implementation of random effects analyses in the context of statistical parametric mapping (Holmes *et al.*, 1998). In this paper we focus on the how conjunction analyses, employing fixed-effect models, can be used to make population inferences about qualitative responses (e.g., activated or not) in terms of confidence intervals for the proportion of the population showing an effect. The proposal here is not to substitute conjunction analyses for random-effect analyses. Where the latter are indicated there is no alternative. This paper simply presents a formal development of the intuition that a careful characterization of a small number of subjects, in a case-study sense, is still useful in establishing typical aspects of functional anatomy. The use of conjunctions provides for a conservative approach to case studies that makes those findings, which may generalize, explicit.

## SOURCES OF SESSION-BY-CONTRAST INTERACTIONS

If one wants to make inferences about a group of subjects the presence of random effects (i.e., differential responses or subject-by-condition interactions) must be considered. The variations in response from one subject to another represent subject-by-condition interactions or, more generally, session-by-contrast interactions (referred to colloquially as “session effects”). They are prevalent in fMRI and their characterization represents one of the current challenges in methodological research. Session effects are important because (i) they violate the assumptions of conventional fixed-effect analyses and (ii) they confound differential activations elicited by the experimental design (e.g., before and after a drug, under different cognitive sets, or in different groups of subjects).

From now on we refer to session-by-contrast interactions, or session effects, and note that subject-by-condition interactions are a special case in which a different subject is scanned in each session and the contrast specifies a condition-specific activation. Session effects can be divided into (i) physiological differences in evoked responses and (ii) instrumentation-dependent differences in sensitivity when measuring those responses. Differences in evoked responses can be intrinsic to a subject or reflect differences among subjects. Within-subject differences can be further divided into task-specific (e.g., adaptation, learning, and strategic changes in cognitive or sensorimotor processing) and nonspecific [e.g., variations in  $p\text{CO}_2$  secondary to hyperventilation or global perfusion changes secondary to changes in antidiuretic hormone secretion in the supine position]. Changes in sensitivity to a given blood oxygenation level-dependent effect can be attributable to many factors (e.g., radio-frequency and gradient instabilities, recalibration of the scanner, repositioning effects, or differential shimming effects) and to interactions among all these factors. In short there are many reasons to suppose that the activation measured in one session could differ from that measured in another. Many of these effects can be minimized or accounted for in terms of *post hoc* modeling; however, their potential to introduce session-by-contrast interactions makes it important to ensure the validity of inferences based on fixed-effect analyses by accommodating them explicitly.

## CONJUNCTION ANALYSES AND SESSION-BY-CONTRAST INTERACTIONS

In Price and Friston (1997) we introduced a simple technique that was designed to identify context-insensitive activations in a series of cognitive subtractions, in other words, a technique that could identify where several activations, in a series of subtractions,

each performed in a different context, were jointly significant (and not significantly different). The essence of this approach is to test for the conjunction of several hypotheses (each specified by a contrast) by assessing the significance of the combined contrasts and then eliminating voxels that evidence significant differences (i.e., interactions) among the contrasts.

The present application of conjunction analysis is motivated from a slightly different perspective and employs some new results (see Worsley, 1999) that provide corrected  $P$  values for conjunction analyses. In the current application “jointly significant” implies that each contrast survives some common threshold. The approach is, however, predicated on the same basic notion of jointly refuting a set of null hypotheses. The motivation for conjunction analyses here is that they allow for (i) an inference at the level of the fixed-effect analysis based on the null hypotheses of no activation in any of the subjects studied and (ii) an inference at a second level, about the population, in terms of a confidence interval for the proportion of the population that is likely to show the regionally specific effect identified at the first level. This second level inference is now considered.

If, for any given contrast, one can establish a conjunction of effects over  $n$  subjects using a test with a specificity of  $1 - \alpha$  and sensitivity  $\beta$ , the probability, under the null hypotheses, of this occurring by chance is

$$P(n) = \sum_{k=0}^n P(n | k) \cdot P(k) = \sum_{k=0}^n \binom{n}{k} (1 - \gamma)^k \gamma^{n-k} \alpha^k \beta^{n-k} \quad (1)$$

$$= [\alpha (1 - \gamma) + \beta \gamma]^n,$$

where  $k$  is the number of subjects sampled that actually possess the effect and  $\gamma$  is the proportion of the population (from which the subjects were sampled) with the effect. Here  $P(n)$  is the probability that  $n$  subjects test positive and is simply the sum of the probabilities of this happening for each possible selection in which  $k$  subjects actually have the effect. In turn this is given by the binomial expansion above, where the probability that a subject sampled at random shows the effect is  $\gamma$  and the probability of a positive test is  $\alpha$  or  $\beta$  given the subject does or does not show the effect, respectively.  $\alpha$  is the  $P$  value of the test and  $\beta$  is its power or sensitivity.<sup>1</sup>

<sup>1</sup> It is of course possible to derive the probability of a conjunction in which  $n$  is less than the number of subjects studied (e.g., the probability of obtaining a positive test in four, or more, of six subjects). However, we do not pursue this here because the corresponding developments in Gaussian field theory, required for corrected inferences, deal only with the situation in which all subjects test positive (see text).-



Now the probability above has an upper bound, in relation to some critical proportion  $\gamma_c$ , that is realized when (the generally unknown)  $\beta = 1$ ,

$$P(n | \gamma \leq \gamma_c) \leq [\alpha(1 - \gamma_c) + \gamma_c]^n. \quad (2)$$

$P(n | \gamma \leq \gamma_c)$  is effectively a  $P$  value corresponding to an inference about how typical the effect is. This inference is at a meta-analytical level and pertains to the population. This second-level inference is based on the classification of subjects (as being significant or not) using a first level of inference pertaining to each subject. In other words, under the null hypothesis [at the second level] that the proportion of the population evidencing this effect is less than or equal to  $\gamma_c$ , the probability of getting a conjunction over  $n$  subjects is equal to, or less than, the left-hand side of Eq. (2). By setting  $P(n | \gamma \leq \gamma_c) = \alpha_c$  (where  $\alpha_c > \alpha$ ) one obtains an expression for this critical proportion,

$$\gamma_c \geq \frac{\alpha_c^{1/n} - \alpha}{1 - \alpha} \quad (3)$$

which approaches unity as  $n$  gets large.  $\alpha_c$  corresponds to specificity at the population level. Its relationship to  $\gamma_c$  is depicted in the top of Fig. 1, which plots the  $P$  values in Eq. (1) against the proportion of the population showing the effect.

In short a conjunction allows one to say, with a specificity of  $1 - \alpha_c$ , that more than  $\gamma_c$  of the population shows the effect in question. Formally, we can view this analysis as a conservative  $100(1 - \alpha_c)\%$  confidence region for the unknown parameter  $\gamma$ . Specifically, the confidence region is  $\gamma > \gamma_c$  if the conjunction occurs, and all values of  $\gamma$  if it does not. It can be checked that the coverage probability, i.e., the probability that the confidence region contains  $\gamma$ , is at least  $(1 - \alpha_c)$ . This approach retains the sensitivity of fixed-effect analyses yet still serves inferences about the population in terms of the proportion affected. These inferences can be construed as statements about how typical the effect is without saying that it is necessarily present in every subject.

### POPULATION LEVEL INFERENCES IN THE CONTEXT OF SPM

The above arguments pertain to the analysis of a single test result (i.e., a single voxel). In a SPM analysis the inference may have to accommodate the fact that one does not know in advance where the effect in question will arise. In this section we consider how a correction for the search volume enters into the inference about the proportion of individuals that are likely to show that effect.

A conjunction analysis of a multisubject/session fMRI study comprises the following steps:

(i) A design matrix is constructed in which the explanatory variables pertaining to each experimental condition are replicated for each session. This subject-separable design matrix implicitly models session-by-condition interactions (i.e., different condition-specific responses among sessions).

(ii) Contrasts are then specified that test for the effect of interest in each session to obtain a set of SPM[ $T$ ].

(iii) These SPM( $T$ ) are thresholded at  $u$  (corresponding to some uncorrected specificity  $\alpha$ ) and combined to give the conjunction.<sup>2</sup> If this combination is on a voxel-by-voxel basis a new SPM that tests for the conjunction is created. The ensuing SPM has two equivalent interpretations: First it represents the intersection of the excursion sets, defined by the threshold  $u$ , of the subject-specific SPM[ $T$ ]. Second it is an SPM of the minimum value of the  $T$  values, thresholded at  $u$ , i.e., a SPM[ $T_{\min}$ ].

(iv) The corrected (for search volume) and uncorrected  $P$  values associated with each voxel in SPM[ $T_{\min}$ ] are now computed as described in the Appendix. These  $p$  values correspond to the null hypotheses of no activation in any subject in which the uncorrected  $P$  value is simply  $\alpha_{\min}^n$ , where  $\alpha_{\min}$  is the specificity given by the threshold  $T_{\min}$ . The corrected  $P$  value  $P_n$  is based on the new results of Worsley (1999; see Appendix). For tabular reporting of significant maxima we use the  $Z$ -variate equivalent of the uncorrected  $P$  value. These corrected or uncorrected  $P$  values provide for inferences about the particular subjects studied based on a subject-separable fixed-effect analysis (the first level). However, because we have demonstrated regionally specific conjunctions, we can now proceed to make an inference (at a second level) about the population from which these subjects came.

<sup>2</sup> This combination can be either in terms of the minimum  $T$  value, over the set of contrasts, at each voxel ( $T_{\min}$ ) or in terms of the minimum of the maximum  $T$  values in some volume of each SPM[ $T$ ] ( $T_{\max}$ ). The former is a special case of the latter in which the volume reduces to one voxel and is the case considered in the main text. Inferences about the ensuing conjunction are then based on the probability of obtaining a maximum  $T_{\min}$  in a search volume of  $S$ . The second case, considered here, is when the minimum of  $T_{\max}$  is used as a statistic: For example, say one expected hippocampal responses in every subject but wanted to allow for individual variations in functional anatomy within the hippocampus. One could take the maximum  $T$  values in a prespecified hippocampal volume and use these  $T_{\max}$  to make an inference at the population level. In this instance the specificity  $\alpha$  in Eq. (2) becomes the probability that any individual  $T_{\max}$  exceeds the threshold, in the specified volume, and is given by  $P_n$  in the appendix, where  $n = 1$ , i.e.,  $P_1$ . The critical proportion is simply given by [cf. Eq. (3)]

$$\gamma_c \geq \frac{\alpha_c^{1/n} - P_1}{1 - P_1}.$$

(v) For uncorrected or anatomically specified inferences the minimum proportion of the population implicated is simply

$$\gamma_c \geq \frac{\alpha_c^{1/n} - \alpha_{\min}}{1 - \alpha_{\min}}, \quad (4)$$

by Eq. (3). Note that for this proportion to be substantially bigger than zero the uncorrected  $P$  value must be less than  $\alpha_c$ . This will, of course, always be the case when  $n$  is greater than one. To make inferences at the population level, which are corrected for the volume of brain analyzed, the situation is a little more complicated and is based on the probability of getting a conjunction anywhere in the brain. In this instance the corrected  $P$  values  $P_n$ , at the subject level, enter as follows. The outcome space, subtending a conjunction over the entire brain, has two partitions, a conjunction in some region in which  $\gamma$  of the population truly show this effect and a chance conjunction elsewhere. The probability of finding a conjunction anywhere is simply one minus the probability that neither of these outcomes occurs. The probability of a positive outcome in the true region of the effect (assuming this region is small<sup>3</sup>) is  $P(n)$ . The probability of a conjunction elsewhere is  $P_n$ , giving

$$\begin{aligned} P(n)_{\text{corrected}} &= 1 - (1 - P(n))(1 - P_n) \\ &= P(n)(1 - P_n) + P_n \geq P(n). \end{aligned} \quad (5)$$

The bottom of Fig. 1 shows examples of  $P(n)_{\text{corrected}}$  using the same threshold, range of population proportions, and sensitivities as in the top (the equivalent probabilities for a single voxel). It can be seen that the correction at the population level corresponds to a slight reduction in the critical proportion if, and only if, the chance probability of a conjunction somewhere ( $P_n$ ) is small. By substituting the corrected probability from Eq. (5) into Eq. (2) the critical proportion is [cf. Eq. (4)]

$$\gamma_c \geq \frac{\left(\frac{\alpha_c - P_n}{1 - P_n}\right)^{1/n} - \alpha_{\min}}{1 - \alpha_{\min}}. \quad (6)$$

<sup>3</sup> A more conservative assumption is that the true effect is expressed over a large region. In the limit of the largest region possible (i.e., the search region)  $P(n)_{\text{corrected}}$  becomes the first component of

$$[\mathbf{A}(1 - \gamma) + \mathbf{I}\beta\gamma]^n \mathbf{b},$$

where  $\mathbf{A}$  and  $\mathbf{b}$  are defined in the Appendix. Although interesting for its symmetry with Eq. (1), this expression is based on unnecessarily conservative and biologically implausible assumptions about the true effect in the sense that regionally specific effect could occur “anywhere” from subject to subject.

In summary the conjunction of a series of session-specific contrasts represents the conjunction of significant effects among several sessions or subjects. In other words they identify activations that are common to all the sessions, in which these activations are inferred on the basis of a fixed-effects model. The presence of a conjunction can be used to infer that at least some proportion of the population from which the subjects came shows this effect. This inference about the population is made on the basis of establishing a confidence interval for the proportion of subjects affected, given a conjunction.

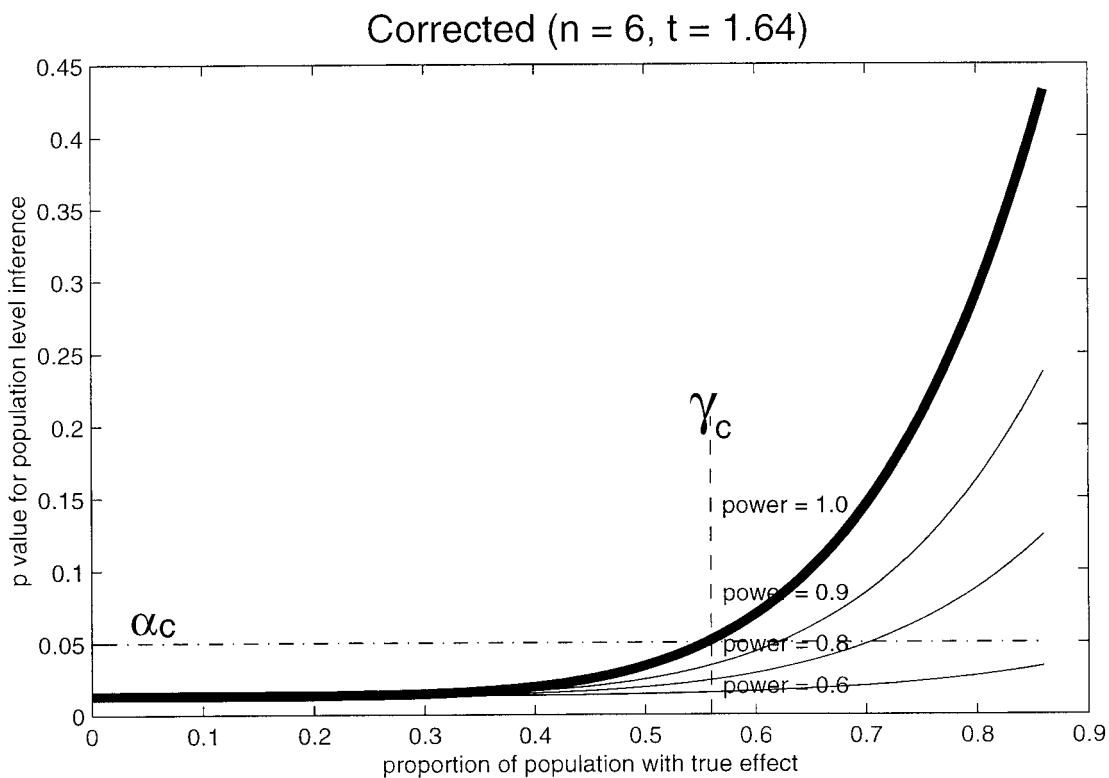
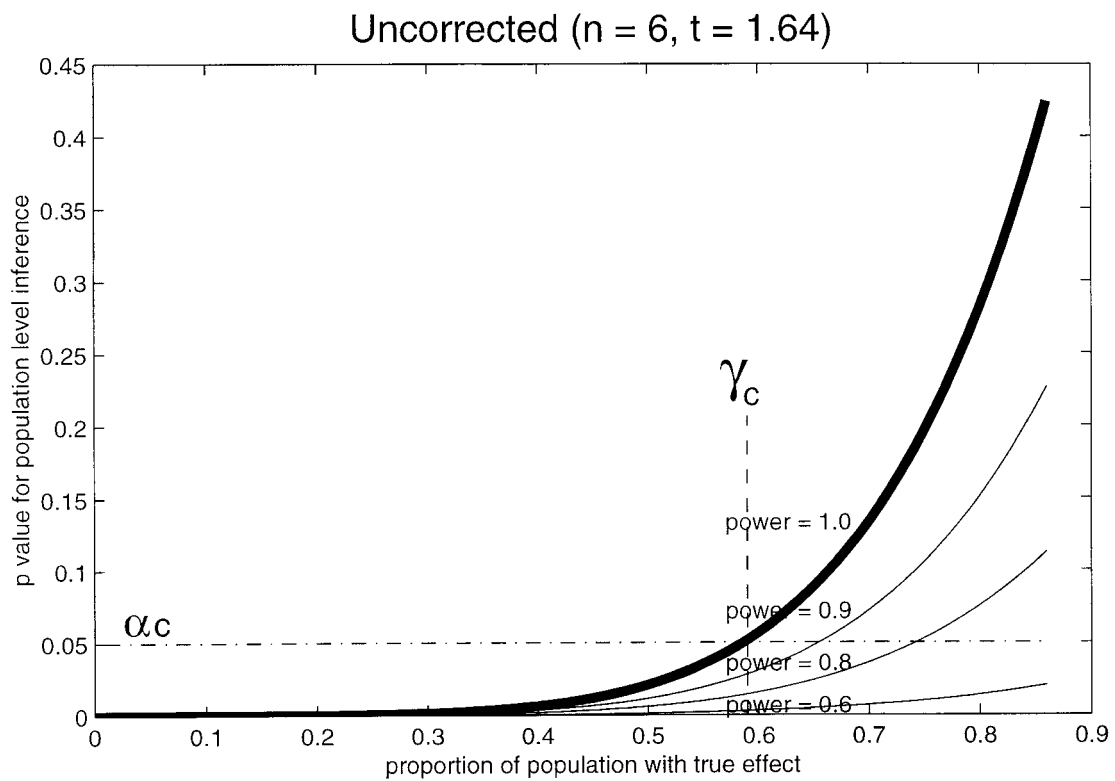
## AN fMRI EXPERIMENT

### Experimental Design and Data Acquisition

The experiment was performed on a 2-T Magnetom VISION (Siemens, Erlangen, Germany) whole-body MRI system equipped with a head volume coil. Contiguous multislice  $T_2^*$  weighted fMRI images (TE = 40 ms; 90 ms/image;  $64 \times 64 \times 32$  3-mm<sup>3</sup> voxels) were obtained with echoplanar imaging using an axial slice orientation. The effective repetition time was 3.22 s. Six young right-handed subjects were scanned under four different conditions, each presented in 10-scan blocks. Two hundred image volumes were acquired for each subject. During all conditions the subjects looked at a fixation point in the middle of a screen. In alternate conditions subjects viewed radially moving dots back-projected on the screen by an LCD video projector. In conditions with visual motion 250 white dots moved radially from the fixation point in random directions toward the border of the screen, at 4.7° per second. The four experimental conditions were “fixation,” “attention,” “no attention,” and “stationary.” In the “attention” condition the subject was asked to “detect changes” in the speed of the dots (which did not actually occur) and during the “no attention” condition the subjects were instructed to “just look.” The fourth condition, “stationary,” consisted of 250 stationary dots. The first few scans of each session were discarded to eliminate magnetic saturation effects.

### Data Preprocessing

The data were analyzed with SPM96 (Wellcome Department of Cognitive Neurology, <http://www.fil.ion.ucl.ac.uk/spm>). The time series were realigned, corrected for movement-related effects, and spatially normalized into the standard space of Talairach and Tournoux (1988) using the subject’s coregistered structural  $T_1$  scan and nonlinear deformations (Friston *et al.*, 1995c, 1996). The data were spatially smoothed with a 6-mm isotropic Gaussian kernel and temporally smoothed with a 6-s Gaussian kernel.



**FIG. 1.** Plots of the probability of obtaining a conjunction, conditional on a certain proportion of the population expressing the effect in question. Top: Plot of  $P(n)$  from Eq. (1) for a test with specificity around 0.05, corresponding to a Z-variate threshold of 1.64, at several sensitivities (1, 0.9, 0.8, and 0.6). The critical specificity for population inferences  $\alpha_c$  and the associated proportion of the population  $\gamma_c$  are

### Statistical Analysis

The contrasts specified compared the visual motion conditions (“attention” and “no attention”) to “fixation.” The design matrix and contrasts used are shown in Fig. 2 (upper right). A design matrix is simply a matrix that contains the explanatory variables in its columns. A linear compound of these columns that best approximates the observed time series in a least squares sense is determined. The coefficients of this compound are the parameter estimates. Effects of interest are specified as a linear compound or contrast of the parameter estimates. Note that the explanatory variables (condition-specific stimulus functions convolved with an estimate of the hemodynamic response function) are replicated separately for each subject, therein modeling all subject-by-condition interactions. In Fig. 2 the contrasts are shown above the design matrix, in register with their associated explanatory variables. For each subject these contrasts were  $[-2, 1, 1, 0]$  pertaining to “fixation,” “attention,” “no attention,” and “stationary,” respectively. Furthermore note that subject-specific effects are also included on the right of the design matrix. These represent another source of random effects and are generally treated as confounds. Other confounds in this design include low-frequency artifacts and global activity.

The ensuing SPM $[T_{\min}]$  testing for a significant conjunction of evoked responses in these subjects is shown in Fig. 2 (top left). All the voxels shown in this maximum-intensity projection expressed a significant ( $P < 0.05$  uncorrected) activation in each of the subject-specific contrasts. It is pleasing to note that both lateral geniculate nuclei (LGN) were significant in this conjunction analysis at a corrected level. The bottom of Fig. 2 shows significant voxels rendered onto a  $T_1$  MRI scan. The locations of the associated LGN maxima were 24,  $-30, 3$  mm ( $Z = 6.62, P < 0.001$  corrected) and  $-24, -27, 0$  mm ( $Z = 8.01, P < 0.001$  corrected). The cross hairs are over the right LGN. At the second level of inference, according the Eq. (4), we can infer with a specificity of 0.05 that more than 60.6% of the population will activate the left LGN under these circumstances. This proportion is very conservative in that it assumes that we are able to detect any such activation with 100% sensitivity (see Fig. 1).

The contrast of parameter estimates corresponding to visual activation are shown in the top of Fig. 3 for the left LGN and a voxel in right V1/V2 ( $-6, -96, 6$  mm). The units are dimensionless and correspond to percentage of the whole-brain mean. The raw (adjusted) and

fitted responses in the V1/V2 voxel are shown in the bottom of Fig. 3 to demonstrate the reproducibility of the response over subjects. This degree of consistency (i.e., absence of session effects) is largely attributable to the conjunction analysis. Consider the example time course in Fig. 4. Here the second subject clearly dominates the responses and there is no evidence for a conjunction. This voxel was identified using a contrast that tested for an average activation over subjects (upper right) and represents a characterization of group responses using a fixed-effect analysis without conjunctions. This voxel came from the right posterior inferior frontal gyrus (48, 6, 24 mm) and had a  $Z$  score of 8.13 ( $P < 0.001$  corrected). It is important to note that this region may still be “activated” according to the contrast used; however, in this instance a fixed-effect analysis is not appropriate for population inference and one would have to use a random-effect analysis.

### DISCUSSION

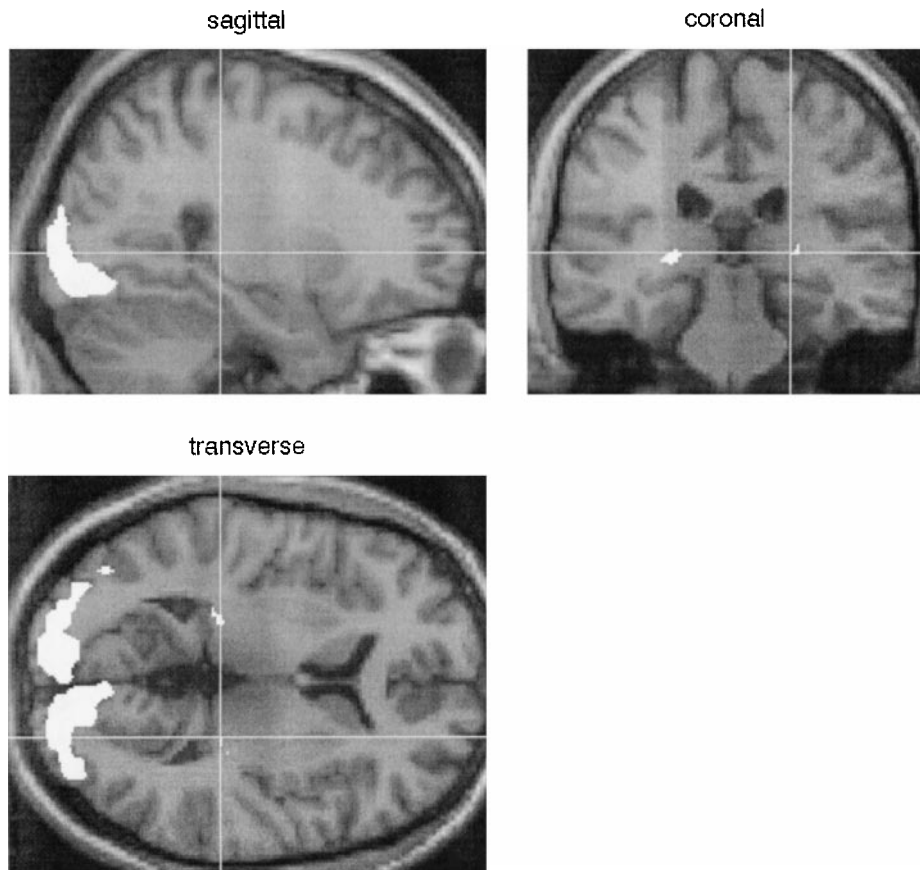
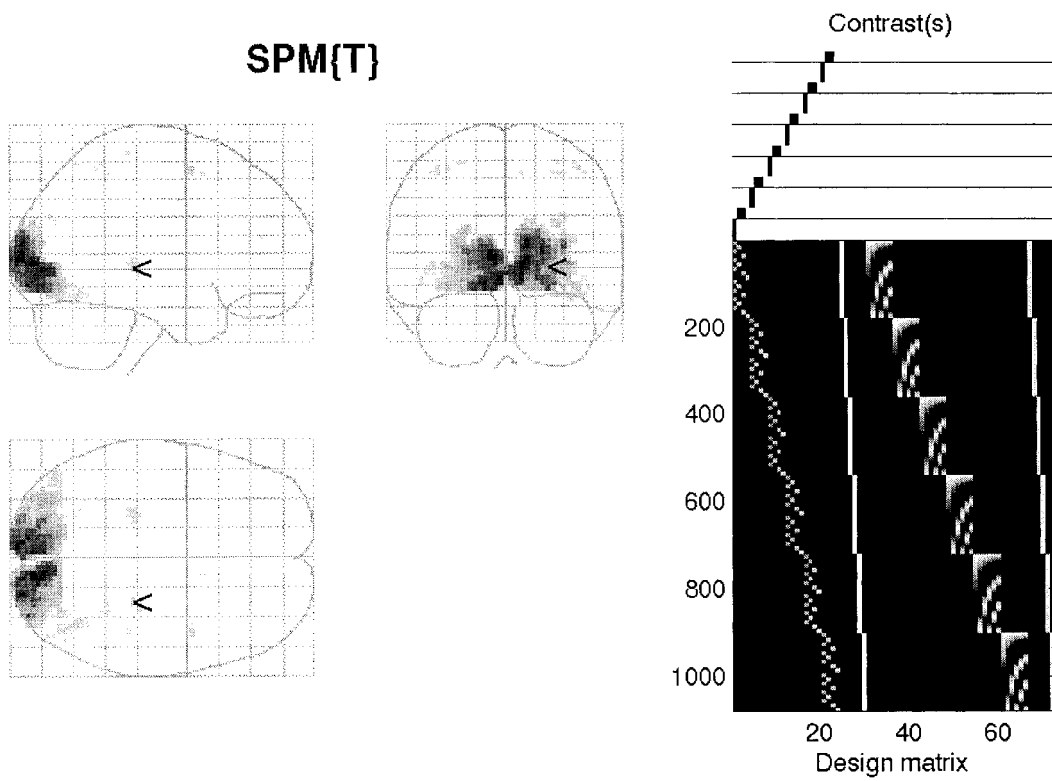
We have presented one approach to the problem of making inferences about activations in multisubject or multisession single-subject designs with fMRI. Activations common to all the subjects studied reflect aspects of functional anatomy that may be “typical” of the population from which the group was sampled. These commonalities can be identified with a conjunction analysis. The motivation behind this use of conjunctions is that fixed-effect analyses are generally more sensitive than equivalent random-effect analyses. The price one pays for the increased sensitivity of fixed-effect analyses is that the ensuing inferences pertain to, and only to, the subjects studied. However a conjunction analysis, using a fixed-effect model, allows one to infer: (i) that every subject studied activated and (ii) that at least a certain proportion of the population would have shown this effect.

The relationship between conjunction analyses, of the sort proposed here, and random-effect analyses needs to be pursued. However, any comparison will be somewhat complicated by the fact that the questions asked by the two approaches are framed in different ways. The conjunction approach effectively finds the confidence region for the proportion of the population showing a categorical effect, whereas the random-effect analysis asks whether the effect is significantly different from zero in quantitative terms. The critical difference is that conjunction analyses allow some subjects not to show an effect and only requires that a substan-

---

denoted by the broken lines. Bottom: As for the top, but in this case the conditional probability  $P(n)_{\text{corrected}}$  is corrected for a typical search volume (that of the fMRI example used in this paper) of 45,415 voxels with a FWHM smoothness of 2.6 voxels (i.e., 2705 resels) in a SPM $[Z]$ . This corresponds to  $P(n)_{\text{corrected}}$  in Eq. (5) where  $P_n$  was computed as described in the Appendix giving  $P_n = 0.0133$  and  $R = [1\ 34.57\ 469.43\ 2705]$ .





**FIG. 2.** Upper right: Design matrix employed in the conjunction analysis. The left-hand partition contains the effects of interest and corresponds to stimulus functions for each of the four photic stimulation conditions convolved with a hemodynamic response function. The four conditions are replicated for each of the six subjects. The order of the conditions is “fixation,” “attention” to visual motion, “no attention” to



tial proportion do. The random-effect analysis posits that the average effect will be bigger than zero. Comparing the sensitivity of the two approaches would therefore involve assessing the power or sensitivity of random-effect analyses, in relation to the probability of detecting a conjunction, for a specified proportion of the population evidencing the effect. This would involve the specification of this proportion, the size of the effect, and the within- and between-subject error variances. This sort of comparison is the subject of future work.

It is important to reiterate that conjunction analyses are not a substitute for random-effect analyses, they are simply an extension of fixed-effect analyses that provide a confidence interval for the proportion of all subjects showing the effect. One application (pointed out by one of the reviewers) is that pilot studies employing conjunction analyses might be a useful prelude to studies of large numbers of subjects that are required by random-effect analyses.

In this paper we have restricted our comments to the analysis of fMRI time series. It is of course possible to apply exactly the same approach to multisubject positron emission tomography (PET) studies. The advantage of doing this is that robust inferences can be made about group effects. Why has this not been an issue for PET? The main reason is that the scan-to-scan variability within a PET session and the session-by-contrast interactions are about the same in PET and the difference between inferences based on fixed- and random-effect analyses is greatly attenuated. Conversely in fMRI the scan-to-scan variability can be much less than the session-to-session differences because (i) fMRI is exquisitely sensitive to small scan-to-scan changes in signal (cf. PET, in which a new bolus of radiotracer is injected at each scan), (ii) there are many more sources of session effects in fMRI (as noted above), and (iii) the scan-to-subject ratio is much higher for fMRI (this ratio dictates the relative weights given to the within- and between-subject variance in random-effect analyses).

It is likely that random-effect analyses will become increasingly important in fMRI, especially when the sphere of inference has to be extended to the population from which subjects were selected (or a subject's general response when scanned on any occasion). A good example of the danger of using fixed-effect analyses is in the study of gender-related differences in functional

anatomy (Price and Friston, 1996). To make an inference that men and women activate differently in response to the same cognitive challenge one has to adopt a random-effect analysis or qualify the validity of the fixed-effect analysis as proposed in this paper. One advantage of the fixed-effect analysis is that, with the use of appropriate contrasts, the subject-by-contrast interactions can be examined explicitly. This may be of interest when trying to characterize intersubject variability. From the perspective of the random effects analysis this variability simply constitutes error.

The proposal here is not to substitute conjunction analyses for random-effect analyses. Where the latter are indicated there is no alternative. This paper is simply an endorsement of the intuition that careful characterizations of small numbers of subjects, in a case-study or fixed-effect analysis sense, are still useful in establishing typical aspects of functional anatomy. The use of conjunctions provides for a conservative approach to these case studies that makes those findings, that may generalize, explicit.

## APPENDIX

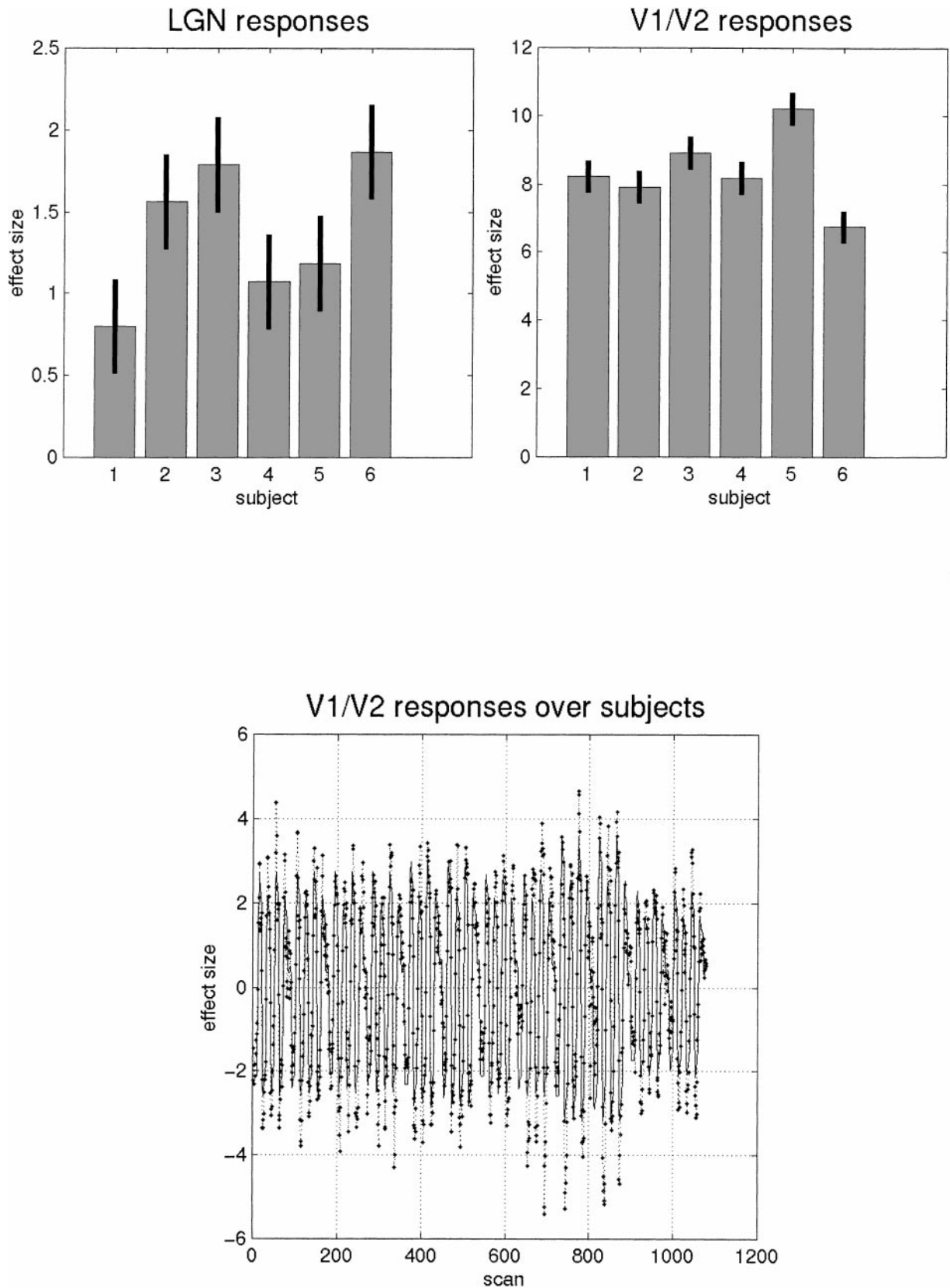
In this Appendix we present the equations that give the corrected  $P$  value  $P_n$  for a conjunction of effects in multiple SPMs of any statistic using (i) the unified theory described in Worsley *et al.* (1996) that can be applied to any specified search volume and (ii) the extension for intersections of multiple SPMs given in Worsley (1999). A conjunction can be modeled as the intersection of the excursion sets  $A_1, \dots, A_n$  of  $n$  isotropic random fields  $X_1, \dots, X_n$  in  $\mathfrak{R}^D$  (or equivalently a single  $D$ -dimensional field whose values comprise the minimum value of  $X_1, \dots, X_n$ ) conjointly thresholded at  $t$ . By the Poisson clumping heuristic

$$P_n = 1 - e^{-\psi_0}, \quad (\text{A.1})$$

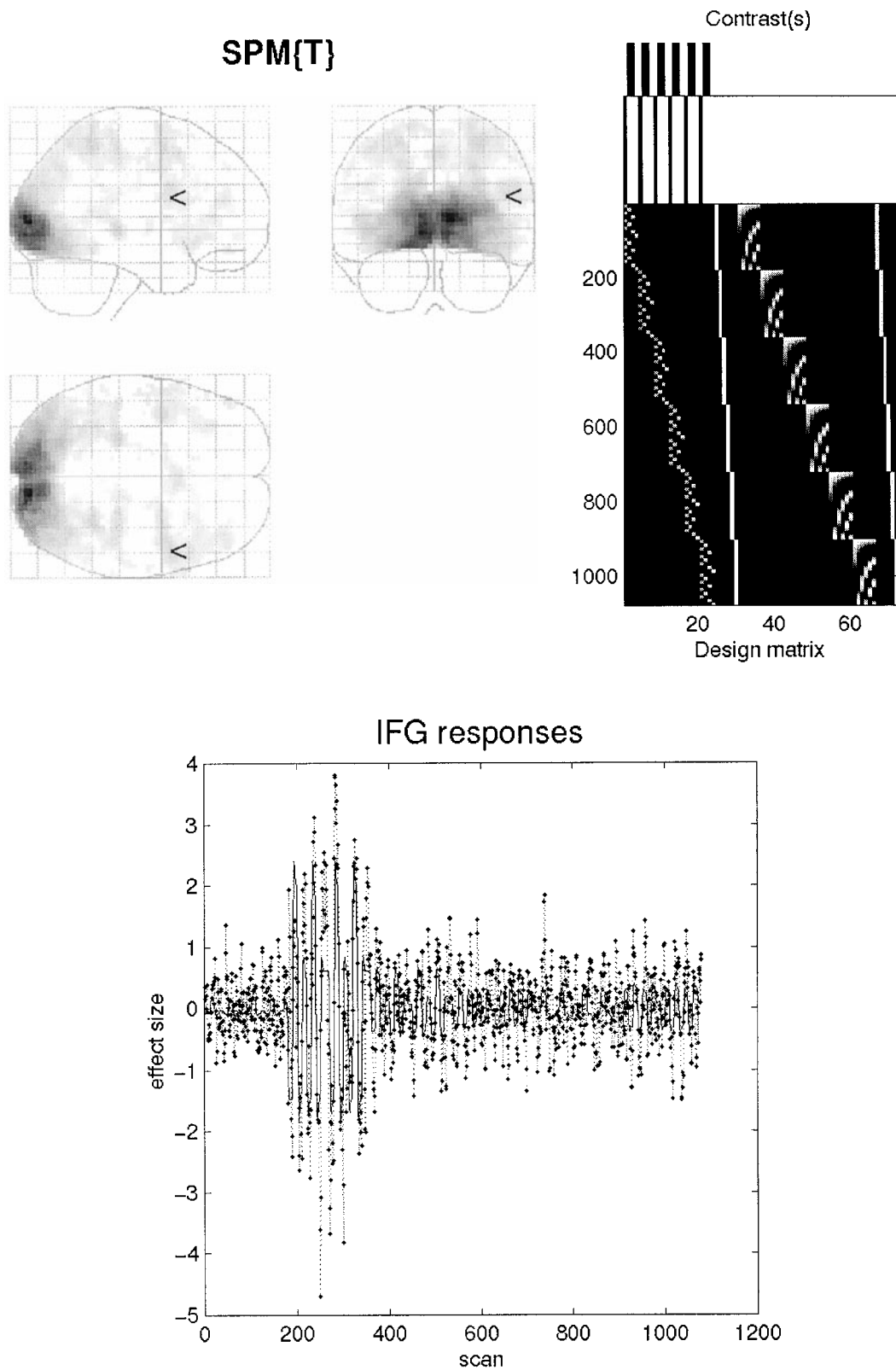
where  $\psi_0$  is an estimate of the expected number of maxima in the conjunction (Adler, 1981; Friston *et al.*, 1994). The expected number of maxima is approximated by the Euler characteristic of the conjunction and is the first component of (Worsley, 1999)

---

visual motion, and "stationary." The remaining partitions (left to right) comprise session or subject-specific effects, session-specific low-frequency confounds (a discrete cosine set), and whole brain activity. The contrasts used are shown above their respective effects in the design matrix. There is one contrast per subject. Upper left: SPM( $T_{\min}$ ) testing for the significance of the conjunction. This is a maximum-intensity projection of a statistical process of the minimum  $T$  statistic over all six subjects. The format is standard and provides three orthogonal projections in an anatomical space conforming to that described in Talairach and Tournoux (1988). The gray scale is arbitrary and the SPM( $T_{\min}$ ) has been thresholded at 1.65 ( $P < 0.05$  uncorrected). Bottom: Significant voxels from the SPM( $T_{\min}$ ) rendered onto a structural MRI scan to demonstrate the location of the LGN responses in this analysis. Both LGN were significant at a corrected level. The maxima of these effects were 24, -30, 3 mm ( $Z = 6.62$ ,  $P < 0.001$  corrected) and -24, -27, 0 mm ( $Z = 8.01$ ,  $P < 0.001$  corrected). The cross hairs are over the right LGN.



**FIG. 3.** Top: Parameter estimates for a voxel in left LGN and V1/V2 ( $-6, -96, 6$  mm) and their standard errors from the fixed-effect analysis. Bottom: Adjusted (dots) and fitted (solid line) responses in the extrastriate voxel plotted over all sessions and subjects (i.e., scans).



**FIG. 4.** Top: As for Fig. 2, but in this instance there is only one contrast testing for an average effect over all subjects. Bottom: As for Fig. 3, but in this instance one subject clearly dominates. The voxel here was at 48, 6, 24 mm ( $Z = 8.13$ ,  $P < 0.001$  corrected).

