

## NIMG 261 REVISION

### Connectivity analysis with Structural Equation Modelling:

#### An example of the effects of voxel selection

Miguel S. Gonçaves\* and Deborah A. Hall<sup>\*CA</sup>

\*MRC Institute of Hearing Research

University Park

Nottingham

United Kingdom

NG7 2RD

<sup>CA</sup> corresponding author

Tel: (+44) 115 922 3431

Fax: (+44) 115 951 8503

email: [debbie@ihr.mrc.ac.uk](mailto:debbie@ihr.mrc.ac.uk)

**Running Title:** Voxel Selection in SEM

**Date of Submission:** 8 November 2002

**Date of revision:** 10 March 2003

**Date accepted:** 28 March 2003

## **Abstract**

Structural equation modelling (SEM) of neuroimaging data is commonly applied to a network of distributed brain regions. We applied SEM to an fMRI dataset to identify condition-specific effects in a simple experiment composed of visual stimulation and baseline conditions. The visual network was composed of three well-defined anatomical regions (V1, V2 and V5) and three path connections (V1→V2, V1→V5 and V2→V5). This network was used to test four hypotheses: (1) whether the condition-specific effects for all three connections vary according to the data selected for modelling; (2) whether the ‘summary’ measures that are often used are indeed appropriate; (3) whether measures taken from the voxel timecourse can reliably predict the condition-specific effects for each one of the three path connections; (4) whether all voxels within an anatomical region yield equivalent SEM outcomes. There was some variability in the significance of the condition-specific effects across randomly-selected voxels within regions. However, the SEM outcome from the ‘summary’ measures was comparable to the most frequent pattern of condition-specific effects. Magnitude, delay, spread and goodness-of-fit measures taken from a gamma fit to the voxel timecourses predicted reliably the significance of the SEM condition-specific effects for each connection. This result enabled us to identify spatially coherent regions at the boundaries of V2 that displayed different condition-specific effects from those seen in the majority of the voxels. Although the generality of these results awaits further investigation, this example highlights a number of important issues for SEM. We have provided further evidence that the SEM outcome does vary somewhat according to the voxels selected and that, while the use of summary measures can give a generalised view of the connectivity pattern, they could fail to capture functional differences within specialised areas.

**Keywords:** fMRI; summary measures; effective connectivity; functional heterogeneity.

## **Introduction**

Multivariate statistical modelling has been applied to neuroimaging data to examine the dynamic functional organisation of the human brain (Grafton et al. 1994; McIntosh and Gonzalez-Lima, 1994; Büchel and Friston, 1997; Jennings et al., 1998; Köhler et al., 1998; Büchel et al., 1999; McIntosh et al., 1998; Fletcher et al., 1999; Bullmore et al., 2000; Della-Maggiore et al., 2000; Gonçalves et al., 2001, Mechelli et al., 2002). Structural equation modelling (SEM) is such a technique that examines covariances between a network of observed and latent dependent variables. When applied to neural systems, the logic of SEM is that brain function is the result of changes in the covariances of activity among different regions that are anatomically interconnected. The pattern of covariances between brain areas is held to be due to common influences and/or the direct anatomical connections between them (McIntosh and Gonzalez-Lima, 1994). SEM is best used in a highly constrained way, typically exploring the cortico-cortical connections between a network of brain regions whose anatomy is explicitly defined by a model. In so far as one brain region may influence the signal within another region, the relationship between regions can be quantified as a covariance. SEM further posits that interactions between regions specified by a model can be detected by examining the covariances of the measured functional signal within the brain (McIntosh and Gonzalez-Lima, 1994). SEM is a powerful statistical technique for neuroscientists because it can determine the significance of condition-specific effects on brain connectivity from the estimates of the path connections in each experimental condition.

Voxels within an anatomical region should lead to similar SEM results, since they are assumed to reflect neuronal tissue that is executing a common cognitive function. Thus, an important issue for SEM methodology is the criterion for data selection. With respect to brain data acquired using functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), different authors have employed different approaches to selection. Some authors identify an

individual voxel within each region to represent the activity within that region. This voxel can be the one with the greatest statistical value, known as the peak voxel of activation, determined by contrasting the experimental conditions (Jennings et al, 1998; Gonçalves et al., 2001). Alternatively, the peak voxel within each region may be determined by a partial least squares analysis in which a ‘seed’ voxel in a key brain region is used as a covariate in a general linear model to identify those voxels in other brain regions of which the time courses are significantly correlated with the ‘seed’ voxel (Horwitz et al., 1995; McIntosh et al., 1998; Della-Maggiore et al., 2000). Other authors select a subset of voxels within each activated brain region. Here, input to the SEM may be the first eigenvariate (from a principal component analysis) of the activity in this group of voxels (Grafton et al. 1994; Büchel and Friston, 1997; Fletcher et al., 1999; Büchel et al, 1999; Bullmore et al., 2000; Honey et al, 2002, Mechelli et al., 2002), or it may be a simple average of those voxels (Köhler et al., 1998). If the image data are spatially smoothed by a broad smoothing function and/or the subset of voxels is small in number, then these measures may not materially differ from the peak voxel.

All of the methods listed for data selection share the assumption that the observed variable provides a reliable and representative summary of the activity within each brain region. To be generalisable, the overall results given by the SEM, such as the fit of the data to the model and the significance of the condition-specific effects on the path connections, should be consistent between data selection methods. The SEM results derived from ‘summary’ variables are only representative if they successfully characterise the majority of results derived using voxels that are randomly-selected within the brain regions. It is well-known that the shape of the stimulus-evoked response varies spatially both within and across brain regions (Lange and Zeger, 1997). If this voxel variation has a functional basis, then the response shape of those selected voxels may influence the pattern of covariance between them, and hence the SEM outcome. If a subset of voxels within a region does not behave in the same way as the majority of the voxels in respect to the SEM results, this may imply the presence of functional differences within the regions. Functional subdivisions may be particularly compelling if these voxels are spatially contiguous.

In this paper, we investigate the effect of the fMRI data selection criterion on the reliability and representativeness of the SEM outcomes for a sample dataset. First, we quantify differences in the condition-specific effects on the path coefficients across a large random sample of voxel datasets; regional heterogeneity. Second, we evaluate whether the ‘summary’ measures often used do indeed correspond to the majority of the results yielded by these randomly selected voxels. Third, using a generalised linear modelling (GLM) procedure, we also assess whether measures of the timecourse of voxel activation contribute significantly to differences in the condition-specific effect. Such measures include the delay, magnitude and duration (spread) of the response. Finally, we identify voxels that yield SEM results differing from the majority of voxels within a selected brain area to determine the spatial pattern of potential regional heterogeneity.

## **Materials and Methods**

### ***fMRI scanning***

Details of the fMRI experiment have been reported in Hall et al. (2002). In summary, a single male volunteer participated in the experiment. He was aged 29, had no history of neurological, hearing or visual impairments and gave informed written consent. MR scanning was carried out using a Varian 3 Tesla system equipped with a fast head gradient coil insert for echo-planar imaging. A scan of 20 T2\*-weighted oblique axial brain images was acquired every 2.5 s, with a 64 by 64 image matrix, a field of view of 22.4 cm, a TE of 30 ms and a flip angle of 90°. The in-plane voxel resolution was 3.5 x 3.5 mm, and the slice thickness was 6 mm. For registration with the volunteer’s own brain anatomy, a 50-slice, T1-weighted structural brain image was obtained at a voxel resolution of 1 x 1 x 3 mm.

Visual stimuli were projected onto a screen placed at the end of the scanner bed and viewed using prism goggles. The visual stimulus was a blue and yellow checkerboard in which the colours

were reversed at a rate of 6.7 Hz. This stimulus was presented for 900 ms and occurred within the period of every 9th scan (i.e. approximately every 22.5 s). Within that period, it could occur at any of 5 different stimulus onset asynchrony values (250, 750, 1250, 1750 and 2250 ms) with respect to the image acquisition. Thus, we were able to reconstruct the time course of the response to the stimulation across a 22.5 s time window with an effective temporal resolution of 500 ms. A total of 695 sets of images were acquired giving 14 data samples for each time point in this window. For an estimation of the baseline signal, 39 sets of images were acquired at the start of the functional imaging run and 16 sets of images at the end.

### ***Image Analysis***

The fMRI data were analysed using SPM99 software (<http://www.fil.ion.ucl.ac.uk/spm>). Each set of brain images was slice-time corrected to the central 10<sup>th</sup> slice and corrected for 3-D head movement. The functional images were normalised into the MNI (Montreal Neurological Institute) brain space using the structural image as a reference. For the purpose of other analysis, not reported here, which required a small final voxel size, normalised images were resampled to 0.8 x 1.0 x 1.9 mm resolution using bi-linear interpolation and the image time series was not spatially or temporally smoothed. Variations in mean global signal intensity were not removed, due to the danger of producing spurious local changes in the direction opposite to any change in global signal. However, to remove aliased respiratory and cardiac effects and other cyclical variations in signal intensity, low-frequency noise was removed using a 0.02 Hz high-pass filter.

An event-related analysis was performed on these data to determine the pattern of visual-evoked activation. The response to each visual event was modelled using a synthetic haemodynamic response function composed of two gamma functions with fixed parameters (to model the haemodynamic response with a slight undershoot) and their temporal derivatives (Josephs et al., 1997). The inclusion of derivatives allowed for some latitude in differential response latencies

across voxels. The T contrast for the response and its temporal derivative was computed at a highly conservative threshold of  $T_{645} > 10$ .

Widespread activation was observed in the occipital cortex including striate (BA 17), parastriate (BA 18) and peristriate (BA 19) cortex. Functionally, this area of activation includes the primary visual region, V1, and non-primary visual regions, V2 and V5. V5 is typically activated by visual motion; and probably here by the dynamic nature of the stimulus. The activation was overlaid onto the corresponding normalised anatomical brain image. Activation was found in both hemispheres in all three visual regions, but for simplicity, we modelled this network for the left hemisphere only. Using anatomical landmarks such as the lingual gyrus and the calcarine sulcus, aided by peak coordinates cited in other studies (Büchel and Friston, 1997; Amunts et al., 2000; Sunaert et al., 2000; Kastner et al., 2001), we defined regions of activation in the left hemisphere for V1, V2 and V5. The boundaries of these regions are shown in Figure 1. On the left, V1 activation encompassed 2,821 resampled voxels, V2 contained 7,892 voxels and V5 contained 2,688 voxels.

**\*\*Figure 1\*\***

### ***Structural Equation Modelling***

Before we could use the voxels in a full SEM analysis, to make inferences about condition-specific effects on the path coefficients, we first need to define which sections of the voxel time series best represent the response to the stimulus and to the baseline conditions. To do this, we used the reconstructed time course of the response to the visual checkerboard stimulus across the 22 s time window. The 45 time points within this experimental cycle were categorised into stimulus, undershoot and baseline intervals (Figure 2). The first 22 time points (0 to 10.5 s) were classified as the stimulation interval and the last 15 time points (15 to and 22 s) were classified as the baseline interval. The central 8 time points (11 to 14.5 s) were classified as the undershoot interval and were

excluded from any further SEM analysis. These selected interval durations may not be ideal for all voxels (e.g. Figure 2 for baseline), but appeared to be reasonable for a majority of voxels. Indeed, prior analysis (reported in Figure 3, Hall et al., 2002) revealed that voxels activated during this paradigm had response delays between 3 and 7 s, thus spanning the stimulation interval defined here. The initial set of 39 images and the last 16 sets of images in the time series were defined as baseline. Thus, the number of data points for each voxel in the SEM was 308 for the stimulation condition and 265 for the baseline condition.

**\*\* Figure 2 \*\***

We defined a network between the three activated visual regions (V1, V2 and V5). Anatomical and functional evidence indicates that V1 and V2 segregate the visual information and parcellate out the different signals to different specialised areas (Zeki, 1993). V1 and V2 are therefore strongly interconnected and both project to V5. Thus, for the application of SEM, the three areas were linked by three connections (V1→V2, V1→V5 and V2→V5) (Figure 1). For parsimony, we do not model feedback connections. The network is not intended to be a realistic model of the visual system, but is a heuristic tool to explore methodological considerations in the application of SEM to fMRI data. We expect from this network that the link between V1→V2 will be particularly affected by visual stimulation, hence we expect a significant condition-specific effect in V1→V2. To investigate this hypothesis, we implemented SEM using a stacked modelling approach described previously (Gonçalves et al., 2001). Here, one level in the stack represented the data for the visual stimulation condition and the other level represented the data for the baseline condition (see Figure 1). Latent variables were applied to the mean adjusted time course of the voxel-based data. For the observed variables, the error term was fixed at 50% of the total variance. The value of the error term determines how much of the variance in the data is to be explained by the model. Given that

variance within a region is unlikely to be accounted for by the connections with the other regions in the model, it should be greater than zero (see McIntosh and Gonzalez-Lima, 1994). Generally, the value of the residual variance ranges from 35-80%. In Gonçalves et al. (2001), we tested four different error values (30%, 40%, 50% and 80%), but found no meaningful changes in the goodness-of-fit of the models or in the path coefficients relative to the 50% value. Thus, in the present SEM, we again fixed the error term at 50 %. The error terms for the latent variables were left unconstrained since there were enough degrees of freedom in the SEM model to estimate them.

We considered two outputs from the SEM; i) the chi-square ( $\chi^2$ ) value, indicating how well the model fitted the observed variance-covariance structure in the dataset, and ii) the path coefficients, indicating how strongly one brain region influenced another. The path coefficients were estimated using bootstrapping methods which defined 95% confidence intervals for the path coefficients (see Gonçalves et al., 2001). In our stacked model, the path coefficients could differ between stimulation and baseline conditions. To assess the presence of any condition-specific effect on the path coefficients, we computed a chi-square difference ( $\chi^2_{\text{diff}}$ ) for every path connection modelled. The  $\chi^2_{\text{diff}}$  represents the difference in the  $\chi^2$  value between two types of SEM models, a constrained model and an unconstrained model. In the constrained SEM model, the path coefficients for the path connection being tested were forced to be equal across both levels. The  $\chi^2$  value for the constrained model indicates how good the model is when one constraint of equality is placed in the path connection of interest. In the unconstrained SEM model, all three path connections were allowed to have different path coefficients across the two levels. This model, in our case, was saturated, i.e. it had zero degrees of freedom, and so the fit of the model to the data was perfect, hence  $\chi^2 = 0$ . A  $\chi^2_{\text{diff}}$  exceeding the value 3.841 (equivalent to  $P < 0.05$ ) indicates that the strength of the corresponding connection significantly differs between the stimulus and baseline conditions. Thus, for example, to assess the condition-specific effect on the connection  $V1 \rightarrow V2$ , we calculated the  $\chi^2_{\text{diff}}$  between the constrained SEM in which  $V1 \rightarrow V2$  was the same across conditions and the unconstrained model in which  $V1 \rightarrow V2$  differed across conditions. This procedure was repeated for

the connections  $V1 \rightarrow V5$  and  $V2 \rightarrow V5$ ; hence, yielding three  $\chi^2_{\text{diff}}$  values ( $\chi^2_{\text{diff } V1 \rightarrow V2}$ ,  $\chi^2_{\text{diff } V1 \rightarrow V5}$  and  $\chi^2_{\text{diff } V2 \rightarrow V5}$ , respectively).

### ***Predicting the pattern of connectivity from measures of the voxel time course***

*Gamma fitting* We wished to determine whether the characteristics of the time course of the selected voxels exert a predictable influence on the outcome of the SEM, for example, in terms of the stimulus-dependent effects defined by  $\chi^2_{\text{diff } V1 \rightarrow V2}$ ,  $\chi^2_{\text{diff } V1 \rightarrow V5}$  and  $\chi^2_{\text{diff } V2 \rightarrow V5}$ . In the present experiment, the shape of the response, ignoring the undershoot, approximates to a gamma curve (see Figure 2). Thus, measures were taken from a gamma fit to the average time course for each of the selected individual voxels using the procedure described by Hall et al. (2002). Four measures were taken for each voxel from its best gamma curve fit; i) magnitude (the area under the positive portion of the fitted curve), ii) delay (the time in seconds, after stimulus onset, to the mean of the fitted response), iii) spread (the temporal width of the gamma function), and iv) the goodness-of-fit (the fit of the gamma to the average time course, where a low value represents a good fit).

*Generalised Linear Model* To test whether the magnitude, delay, spread and goodness of fit response measures contribute significantly to differences in the condition-specific connectivity, we used a GLM to model the relationship between the response measures and the  $\chi^2_{\text{diff}}$ . A traditional linear regression model is not appropriate in this context because the  $\chi^2_{\text{diff}}$  data are not normally distributed ( $\chi^2_{\text{diff}}$  take only positive values). The GLM extends the traditional linear regression model, so that it is applicable to a wider range of problems. A GLM consists of a link function, which applies a data transformation within the model, and the choice of the link function is dependent on the distribution of the data modelled. The general formula of a GLM is:

$$g\{E(Y)\} = \mathbf{x}\boldsymbol{\beta}$$

where  $g$  is the link function,  $E(Y)$  is the expected value or mean of the dependent variable ( $Y$ ),  $\mathbf{x}$  are the covariates of interest and  $\boldsymbol{\beta}$  are the estimated parameters. The  $\chi^2$  data for the differences is gamma distributed since it is skewed to the right and the natural link function for the gamma distribution is the reciprocal ( $g(x) = 1/x$ ) (McCullagh and Nelder, 1983). However, in this particular case, we chose to use the log link function since it achieved better fitting and stable models. The GLM for modelling the  $\chi^2_{\text{diff}}$  at each connection was defined as,

$$\begin{aligned} \ln\{E(\chi^2_{\text{diff V1}\rightarrow\text{V2}})\} &= \mathbf{x}\boldsymbol{\beta}, & \chi^2_{\text{diff V1}\rightarrow\text{V2}} &\sim \text{Gamma distribution} \\ \ln\{E(\chi^2_{\text{diff V1}\rightarrow\text{V5}})\} &= \mathbf{x}\boldsymbol{\beta}, & \chi^2_{\text{diff V1}\rightarrow\text{V5}} &\sim \text{Gamma distribution} \\ \ln\{E(\chi^2_{\text{diff V2}\rightarrow\text{V5}})\} &= \mathbf{x}\boldsymbol{\beta}, & \chi^2_{\text{diff V2}\rightarrow\text{V5}} &\sim \text{Gamma distribution} \end{aligned}$$

For each GLM, we determined the voxel measures that best predicted the  $\chi^2_{\text{diff}}$  using a backward stepwise approach. To yield GLM parameter estimates of a comparable range for the four voxel measures we standardised the response measures to have mean equal to zero and standard deviation equal to 1. The first model iteration included all four response measures (magnitude, delay, spread and goodness-of-fit) plus the square of magnitudes and delays, and the product of magnitude and delay for the three voxels in V1, V2 and V5 to account for any non-linear structure in the model. The first iteration had 21 covariates in total. The least significant covariates were eliminated sequentially until all the covariates in the GLM had a  $P < 0.2$ .

### ***Voxel selection***

*Individual-voxel datasets* To determine whether the outcomes of SEM were dependent upon the voxels selected within each region, and to derive the optimal set of covariates for each GLM, 120 voxels were selected at random within each region of V1, V2 and V5 (for voxel information and location consult <http://www.ihr.mrc.ac.uk/caf/loctable.pdf>). These voxel datasets were used to generate 120 three-way datasets representing different examples of the network for modelling. For each dataset, the observed  $\chi^2_{\text{diff}}$  were computed from the SEM and the corresponding predicted  $\chi^2_{\text{diff}}$  were computed from the GLM.

*Summary datasets* In order to compare the SEM outcomes for the individual randomly-selected voxel datasets with those for ‘summary’ measures, we computed three commonly-used summary measures for the signal in each region, the peak voxel, the first eigenvariate and the simple average. The voxel with the highest T value defined the peak voxel for each region. Voxels located within a 5 mm radius of each peak were then used to calculate the first eigenvariate and the simple average. The 5 mm radius yielded 348 voxels for V1, 324 for V2 and 297 for V5.

*Additional datasets for testing the GLM* Clearly, the forecasting accuracy of the GLMs would be over-estimated if the GLMs were tested using the same datasets that had been used for their derivation. Therefore, to test the predictive power of each GLM, we compared the observed and predicted  $\chi^2_{\text{diff}}$  values for additional datasets. The test set comprised 12 randomly-selected individual-voxel datasets; 10% of the size of the original 120 datasets. The general predictive power of the GLMs was judged according to whether or not they successfully classified both significant and non-significant condition-specific effects, rather than whether or not they produced the correct absolute  $\chi^2_{\text{diff}}$  value.

Thus, in total, we applied SEM to 135 datasets: 120 randomly-selected individual-voxel datasets, the peak, the first eigenvariate, the simple average and the 12 randomly-selected individual-voxel test sets.

### ***Functional differences within specialised areas***

The logic outlined in the Introduction posited that if a region of interest is functionally homogenous, then the significance of the condition-specific effects for the path connections should be consistent, irrespective of which voxels within that region are sampled for the SEM. The consistency of the SEM observed  $\chi^2_{\text{diff}}$  for the 120 individual-voxel datasets from this subject is one test of this assumption. A further test of the assumption can be made using the GLM predicted  $\chi^2_{\text{diff}}$ . This second test is arguably the more exhaustive of the two. The GLM approach is much more economical in computational terms, and so we can select each and every voxel within a region for modelling, rather than use a restricted sample as would be the case for SEM. Furthermore, models used to generate the GLM predicted  $\chi^2_{\text{diff}}$  will always converge, since they rely solely on voxel measures to attain a result.

To exemplify the approach, we fixed the data representing regions V1 and V5 by using the voxel response measures for the peak voxel in each region. Fixing the voxels in V1 and V5 facilitated the graphical representation of the role of V2, by simply generating two maps to illustrate condition-specificity of the links between V1→V2 and V2→V5 respectively. We then computed GLM outcomes using response measures for every one of the 7892 voxels within the third region, V2. This enabled us to assess how the condition-specific effects vary in V2 when it is mapped onto one voxel from V1 or V5. Given that we varied the data for V2, our assessment focused on the predicted  $\chi^2_{\text{diff}}$  for the two connections that involved V2 (V1→V2 and V2→V5). We chose to vary V2 because this area played the greatest role in the SEM (see Results). For each GLM, the predicted  $\chi^2_{\text{diff}}$  values were dichotomised ( $\geq$  or  $<$  3.841). The distribution of predicted outcomes can be

displayed as a spatial map to assess whether voxels within V2 behave differently from the representative peak voxel and if so, whether these are spatially co-localised.

## Results

*120 individual-voxel datasets* The SEM for 11 of the 120 datasets did not successfully converge after 500 iterations and were eliminated from further analysis. Of the remaining 109 datasets, the unconstrained model gave a wide range of absolute values of the path coefficients for each connection (see Figure 3). However, the main focus of interest for this paper is on the significance of the positive difference between the stimulation and baseline condition (condition-specific effect where stimulation is greater than baseline), rather than on the absolute value of the path coefficient. The models also differed in terms of whether or not a connection showed a significant condition-specific effect (also illustrated in Figure 3). For V1→V2, 66% (72/109) of the models indicated a significant effect of stimulus on the strength of this connection. 27% (29/109) of the models had a significant stimulus-specific effect for V1→V5 and 17% (19/109) of the models showed a significant stimulus-specific effect for V2→V5. Two models for V1→V5 and four models for V2→V5 showed a significant baseline-specific effect. All other models showed non-significant condition-specific effects.

**\*\* Figure 3 \*\***

The response measures derived from the gamma curve fits to the data for all 120 datasets are summarised in Table 1. Across regions, there were some marked differences. The average voxel magnitude and spread were significantly lower in V5 than in V1 and V2 ( $P < 0.001$  and  $P < 0.05$ ,

respectively). The average voxel delay was significantly shorter in V5 than in V1 ( $P < 0.05$ ), but not shorter than that in V2 ( $P > 0.05$ ). The gamma curve fits to the responses in V5 were significantly worse than to those in both V1 and V2 ( $P < 0.001$ ).

**\*\* Table 1\*\***

*Summary datasets* The SEM model failed to converge for the first eigenvariate dataset because the variance of each eigenvariate was very close to zero. The implications of this finding are considered in the Discussion. The SEM model for the peak voxels converged after 12 iterations. The path coefficients for all three connections fell within the range of the initial 120 datasets and their values are reported in Figure 3. V1→V2 showed a significant stimulus-specific effect ( $\chi^2_{\text{diff}} = 18.79$ ,  $P < 0.001$ ). Condition-specific effects were absent for V1→V5 and V2→V5 ( $\chi^2_{\text{diff}} = 0.61$ ,  $P = 0.43$  and  $\chi^2_{\text{diff}} = 0.24$ ,  $P = 0.63$  respectively). The response measures derived from the gamma curves fitted to the data for the peak voxel are summarised in Table 1. The SEM for the average of the 5mm radius sphere around the peaks was equivalent to that of the peak voxel (Figure 3).

In summary, SEM outcomes for both the peak voxels and the average around the peak were consistent with the general outcome of the SEM results from the individual-voxel datasets with respect to the presence of a stimulus-specific effect on the V1→V2 connection.

*Additional 12 individual-voxel datasets for testing the GLM* The SEM for one of the 12 test sets did not successfully converge and was eliminated from further analysis. Figure 3 shows the range of the path coefficients and the significance of the condition-specific effects on the network connections for the test set. For V1→V2 path coefficients were within the range shown by the initial dataset for the connection (Figure 3). However, for V1→V5, one of the additional datasets

had stimulus and baseline coefficients that exceeded those in the initial 120 dataset. In addition, for V2→V5 two of the additional datasets had stimulus and baseline coefficients that were lower than those in the initial 120 datasets. Again, a high proportion of datasets revealed a significant stimulus-specific effect for V1→V2 (82%). For the other two connections, a significant stimulus-specific effect was present for only 9% of V1→V5 connections and for 27% of V2→V5 connections.

The gamma fit measures for the 12 test sets were not significantly different to those for the 120 datasets for both V1 and V2 (Table 1). However, in V5, the average magnitude was statistically smaller ( $P < 0.05$ ) for the 12 test sets.

*Generalised Linear Model* Each GLM, derived from the 120 individual-voxel datasets, described which set of response measures best predicted the estimate for  $\chi^2_{\text{diff V1} \rightarrow \text{V2}}$ ,  $\chi^2_{\text{diff V1} \rightarrow \text{V5}}$  and  $\chi^2_{\text{diff V2} \rightarrow \text{V5}}$  respectively. Each GLM model had 6 covariates, whose significance was  $P < 0.2$ , although not all the same covariates reached significance across the three models. The covariates used in each model are reported in Table 2. The response measures that most contribute to the final GLMs (13 out of the total of 18 covariates) are the magnitude and delay. A greater number of the measures from V2 reached statistical significance ( $P < 0.2$ ) than from V1 and V5, indicating a larger role for V2 in the visual network.

\*\* Table 2 \*\*

For comparison between predicted and observed  $\chi^2_{\text{diff}}$ , we used the inverse link function to linearize the predicted  $\chi^2_{\text{diff}}$  from the GLMs. Figure 4 shows the plots of the observed values for the  $\chi^2_{\text{diff V1} \rightarrow \text{V2}}$ ,  $\chi^2_{\text{diff V1} \rightarrow \text{V5}}$  and  $\chi^2_{\text{diff V2} \rightarrow \text{V5}}$  against the GLM predicted values. For  $\chi^2_{\text{diff V1} \rightarrow \text{V2}}$ , the GLM model correctly identified the significant and non-significant path differences in 80% of the 109

different connections. For  $\chi^2_{\text{diff V1}\rightarrow\text{V5}}$  and  $\chi^2_{\text{diff V2}\rightarrow\text{V5}}$ , the GLM correctly identified 73% and 79% respectively. The predictive power of the GLM was evaluated using the 11 test sets. The GLM correctly identified significant and non-significant  $\chi^2_{\text{diff}}$  values in 9/11 (82%) cases for V1→V2, all 11 cases for V1→V5 (100%) and 7/11 (64%) cases for V2→V5 (Figure 4). Thus, these data suggest that the gamma measures used in the GLM can reliably predict which voxels within a region will yield significant condition-specific effects in the SEM.

\*\* Figure 4 \*\*

### *Functional differences within specialised areas*

Figure 5 shows axial views of V2 indicating voxels that behave differently from the representative peak voxel with V1 and V5 fixed, for V1→V2 and V2→V5 connections. For the connection that projects from V1 to V2, the red voxels along the borders differed in their behaviour in that they showed a non-significant effect of stimulus condition (Figure 5A). For V2→V5, voxels in blue along the edge (slices  $z = 11$  and  $13$  mm) were the only ones to show a significant condition-specific effect (Figure 5B). Thus, we were able to identify sub-regions within V2 in which the V1→V2 and V2→V5 connections displayed different condition-specific effects from the region as a whole.

\*\* Figure 5 \*\*

## **Discussion**

Although the precise nature of the voxel influences on the SEM outcome await further investigation using additional datasets, the example reported here is important in highlighting a number of issues

in the connectivity modelling of fMRI data. In this paper we used data from a single subject to test four important hypotheses concerning the application of SEM to fMRI data. First, we demonstrated a wide variability in the path strengths for both stimulation and baseline conditions for 120 randomly selected datasets and showed that the SEM outcome does vary somewhat according to the voxels selected. Nevertheless, a majority of the voxel datasets provided consistent condition-specific effects in terms of significance level. Second, the significance of the condition-specific effects from the summary measures (peak and average) were consistent with the majority of the 120 randomly-selected voxels. This finding indicates that commonly used voxel summaries are acceptable ways of summarising the overall information, but do risk missing any functional heterogeneity. Third, in this example we have shown that characteristics of the voxel time course (especially magnitude, delay and spread) contribute significantly to the determination of whether or not a condition-specific effect is present in the SEM. Finally, we have demonstrated that SEM can be used to explore functional heterogeneity within an anatomical area. To illustrate this point, we showed that in V2 there were localised subregions at the boundaries of the region where the condition-specific effect behaved differently from that for the majority of the voxels.

### **Contribution of timecourse measures to SEM**

The predictive power of the three GLM models provides evidence of the usefulness of the information within the voxel time course to determine the significance of the condition-specific effects in the SEM. Hence, this approach has the potential to be a useful tool to aid voxel selection. One caveat however, is that those measures contributing the most to the GLM may vary across path connections, and probably also across subjects and different anatomical models. The general value gained by using timecourse measures to inform the SEM requires further investigation. The significance of the condition-specific effect is likely to be determined by the data in the stimulus rather than the baseline condition, since the magnitude, delay and spread measures only correspond to the stimulus condition. Thus, it is important that there be a clear distinction between the

measures for each experimental condition, plus a good signal-to-noise ratio, or the ‘true’ condition-specific effects may be wrongly estimated. In the GLM model, we included an additional measure (the goodness of fit of the gamma curve to the averaged time course) to control for the poorly fitting curves, since these curves will generate less reliable estimators of magnitude, delay and spread. Additionally, the strong relationship between the voxel time course measures and the condition-specific effects can be used as a tool to efficiently and robustly select sets of voxels that behave similarly within an anatomical region. Other statistical tools, such as independent component analysis (McKeown et al., 1998), may carry out such discrimination more efficiently, but the main advantage of the GLM approach is that it also enables us to discriminate time courses which, despite being dissimilar in their characteristics, achieve similar condition-specific effects in the SEM. However, the success of the GLM approach described here is dependent upon the quality of the gamma curve fit to the imaging data. Here the experiment was carried out with the specific purpose of extracting reliable and robust estimates of the time course measures. The fine temporal resolution of 500 ms and the 14 data values for each time point allowed us to fit gamma curves reliably and obtain quality estimators of the time course. In many experimental designs, it is unlikely that such measures from the time course could be estimated as reliably.

In the present study, we used the GLM with gamma errors to identify which measures were able to predict the significance of the  $\chi^2_{\text{diff}}$  values. However, other modelling strategies could have been used. One popular approach is the logistic regression (which is the same GLM with binomial errors). The logistic regression may seem more intuitive, since we were mainly interested in the dichotomy between the significant and non-significant condition-specific effects. However, we chose to use a more generalised approach to incorporate all the information contained in the  $\chi^2$  values. Manifestly, a  $\chi^2$  value of 18 reveals a much more significant condition-specific effect than a  $\chi^2$  value of 4; in the logistic regression model we would have coded both as one, losing important information within the modelling process. This loss of information from logistic regression would have compromised the robustness and, consequently, the predictive power of the model.

## **Appropriateness of summary measures**

We have shown that both the peak voxel and the average of nearby voxels around the peak voxel are reliable summary measures of the condition-specific effects between areas. The significance values of the condition-specific effects for both datasets were in agreement with the majority of the results from the 120 individual-voxel datasets and, most importantly, they were in agreement with the *a priori* expectation that the visual areas V1 and V2 are strongly connected and show a large effect of the visual stimulus. Despite good agreement in terms of the significance of the effect, the strength of the path coefficients encompassed a broad range for both stimulus and baseline conditions. However, in SEM, the principal concern is whether or not the path coefficients show a change across conditions, rather than what is their absolute value (McIntosh and Gonzalez-Lima, 1994).

These findings appear to indicate that voxel selection within the visual cortex is not a major issue when applying the data to SEM. However, this indifference should not be overstated. If voxels are arbitrarily selected, there is a risk of choosing false-negative voxels (i.e. voxels in which the SEM outcome shows no significant condition-specific effect when there really is one). The risk of false-negatives may be even greater in brain regions beyond the visual cortex. In our experience, the visual cortex is strongly responsive in fMRI studies and generates an MR response that has a high signal-to-noise ratio. For brain regions that have poorer signal-to-noise ratio characteristics, the differences between experimental conditions will be blurred. As a consequence, the correlation structure will increasingly depend on the background activity (noise) than on the stimulus-induced signal. In addition, areas in which the anatomical and/or functional borders are ill-defined may generate inconsistencies in SEM outcome, since the SEM reported here provided divergent results close to the border of V2. This last point could also pose a problem for generalisation when a network of connectivity is postulated to be the same across a group of subjects (Gonçalves et al., 2001). In our experiment, we present data for one subject and the brain regions selected were

relatively large and well-defined, so the effects of precise boundary location on the summary measure will be minimal. However, with smaller brain regions or with those whose borders are not distinct, the boundaries will have a larger influence on the overall behaviour of a region and may consequently contaminate the summary measures. Moreover, if one region in a network is very small, relative to the voxel size, data smoothing may confound the true effect of a region by including a contribution from other surrounding regions. All these reasons justify exercising careful control for the effects of voxel selection and subject differences when modelling any brain network.

While the peak of activation and the average of nearby voxels around the peak were successful measures for use in SEM, the first eigenvariate measure did not produce a converging model. The first eigenvariate was obtained using the SPM99 singular value decomposition approach. In singular value decomposition, the variance of the eigenvariate is scaled to be  $1/(n-1)$ , where  $n$  is the number of points in the time series. In our SEM time series where there were 573 data points, the variance value for the observed variables was close to zero. The low variance created a series of convergence problems in the SEM, arising because our model specified both observed and latent variables. The advantage of the latent variable approach is that one can control for measurement error but, for this control to be efficient, the true variance structure needs to be maintained. If the observed variables have a variance that is close to zero, then it is likely that the error term of the latent variables becomes negative (this is known as Heywood cases). Since the error term of the latent variable represents its variance, it cannot take a negative value, and hence the model does not converge. Most authors who apply connectivity modelling to neuroimaging data use only observed (and not latent) variables in a path analysis technique (Grafton et al. 1994; Büchel and Friston, 1997; Fletcher et al., 1999; Büchel et al., 1999; Bullmore et al., 2000; Honey et al., 2002; Mechelli et al., 2002). Therefore, the scaling of the variance of the first eigenvariate poses no problem for model convergence, since only the correlation structure between the observed variables is of interest and no measurement error is taken into account.

A commonly used solution to the low variance problem when using the latent variable modelling approach is to set the error of the observed variables to zero, instead of 50% (Dillon et al. 1987). It has been shown that fixing the error to zero is adequate if the cause of the negative variance is induced by sampling fluctuations (Dillon et al. 1987). However, this procedure suffers from a number of drawbacks in that the maximum likelihood theory has not been proven to be valid at boundary minima and setting the negative error to zero results in the observed variables having no measurement error, which is unlikely in fMRI. An alternative solution is to re-scale the eigenvariates, by regressing the average time series on the first eigenvariate and then scale the eigenvariate by the slope of this regression (Lawley and Maxwell, 1971). We have implemented both these methods and both successfully eliminated the negative variance problems, enabling the SEM to converge. The results of the SEM again revealed a single significant condition-specific effect for V1→V2 ( $P < 0.001$ ). Similar path strengths to those for the peak voxel and the average analyses were also obtained.

### **Functional homogeneity**

The somewhat variable SEM outcome for the 120 randomly selected datasets, may indicate some functional heterogeneity within a region. The variability within a region may be random or systematic. If the variability is random, the voxels in which the condition-specific effects are different from the majority are more likely to be scattered throughout the region. Alternatively, if the variability is systematic, these voxels are more likely to be spatially coherent. For the single subject reported here, we showed that this variability was indeed systematic because the differences occurred only along the edges of the brain region tested (V2). For simplicity, we reported only one view of a multi-dimensional problem because the map displaying the condition-specific effects is dependent on the voxel selection within the fixed regions. In the example reported, we fixed the voxels in V1 and V5 to be the peak of activation, and then measured the effects across all voxels in V2. A change in the selected voxels within either V1 or V5 will influence the outcome of the SEM,

thus changing the map of V2. Nevertheless, if the peak voxel is an appropriate summary measure for SEM, then the technique that we have described may be informative for data exploration. We speculate that SEM may be a useful technique for identifying functional subdivisions within *a priori* anatomically-defined regions, particularly when testing hypothesis for networks where little is known about their function. The visual cortex perhaps presents a special case where its functional organisation is well established. Retinotopic mapping has demonstrated a reliable relationship between the cortical surface and the visual field whereby mapping responses to the eccentricity and polar angle of a visual stimulus can be used to delineate functional boundaries between visual fields (see Warrington et al., 2002 for a review of the retinotopic mapping technique). A greater degree of uniformity in terms of the SEM outcome for different selected voxels might be achieved if the visual regions were defined using these functional criteria, although this proposition awaits investigation.

## **Conclusion**

This example has demonstrated that aspects of connectivity obtained from SEM can be reliably predicted using measures from the voxel time course such as magnitude, delay and spread. The use of the peak of activation voxel to model the network in SEM gives consistent results with the most other voxel selection strategies. However, in regions where voxel response characteristics are highly variable, problems of homogeneity will occur, interpretable as localised functional subdivisions. Consequently, SEM may be a useful technique for identifying functional differences within specialised areas.

## **Acknowledgements**

We thank Mark Haggard for helpful discussion about the manuscript, Paul Matthews for providing access to the 3 Tesla MR scanner at the Oxford Centre for Functional Magnetic Resonance Imaging of the Brain and Cameron Shaw for assistance with the art work.

## References

- Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T. and Zilles, K. 2000. Brodmann's areas 17 and 18 brought into stereotaxic space - Where and how variable? *NeuroImage* **11**: 66-84.
- Büchel, C. and Friston, K. J. 1997. Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex* **7**: 768-778.
- Büchel, C., Coull, J. T. and Friston, K. J. 1999. The predictive value of changes in effective connectivity for human learning. *Science* **283**: 1538-1541.
- Bullmore, E., Horwitz, B., Honey, G., Brammer, M., Williams, S. and Sharma, T. 2000. How Good Is Good Enough in Path Analysis of fMRI Data? *NeuroImage* **11**: 289-301.
- Della-Maggiore, V., Sekuler, A., Grady, C. L., Bennett, P. J., Sekuler, R., McIntosh, A. R. 2000. Corticolimbic Interactions Associated with Performance on a Short-Term Memory Task Are Modified by Age. *The Journal of Neuroscience* **20**: 8410–8416.
- Dillon, W. R., Kumar, A. and Mulani, N. 1987. Offending Estimates in Covariance Structure Analysis: Comments on the Causes of and Solutions to Heywood Cases. *Psychological Bulletin* **101**: 126-135.
- Fletcher, P., Büchel, C., Josephs, O., Friston, K., Dolan, R. 1999. Learning-related Neuronal Responses in Prefrontal Cortex Studied with Functional Neuroimaging. *Cerebral Cortex* **9**: 168-178.
- Gonçalves, M. S., Hall, D., Johnsrude, I. S., Haggard, M. P. 2001: Can meaningful effective connectivities be obtained between auditory cortical regions?. *NeuroImage* **14**: 1353-1360.

- Grafton, S. T., Sutton, J., Couldwell, W., Lew, M. and Waters, C. 1994. Network analysis of motor system connectivity in Parkinsonis disease: modulation of thalamocortical interactions after pallidotomy. *Human Brain Mapping* **2**: 45-55.
- Hall, D. A., Gonçalves, M. S., Smith, S., Jezzard, P., Haggard, M. P., Kornak, J. 2002. A method for determining venous contribution to BOLD contrast sensory activation. *Magnetic Resonance Imaging*. In press
- Honey, G. D., Fu, C. H. Y, Kim, J., Brammer, M. J., Croudace, T. J., Suckling, J., Pich, E. M., Williams, S. C. R. and Bullmore, E. T. 2002. Effects of Verbal Working Memory Load on Corticocortical Connectivity Modeled by Path Analysis of Functional Magnetic Resonance Imaging Data. *NeuroImage* **17**: 573-582.
- Horwitz, B., McIntosh, A. R., Haxby, J. V., Furey, M., Salerno, J. A., Schapiro, M. B., Rapoport, S. I. and Grady, C. L. 1995: Network analysis of PET-mapped visual pathways in Alzheimer type dementia. *NeuroReport* **6**: 2287-2292
- Jennings, J. M., McIntosh, A. R. and Kapur, S. 1998: Mapping neural interactivity onto regional activity: An analysis of semantic processing and response mode interactions. *Neuroimage* **7**: 244-254.
- Josephs, O., Turner, R., and Friston, K. J. 1997. Event related fMRI. *Human Brain Mapping*, **5**: 243-248.
- Kastner, S., De Weerd, P., Pinsk, M. A., Elizondo, M. I., Desimone, R., Ungerleider, L. G. 2001. Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *Journal of Neurophysiology* **86**: 1398-1411.
- Kohler, S., McIntosh, A. R., Moscovitch, M., Winocur, G. 1998. Functional interactions between the medial temporal lobes and posterior neocortex related to episodic memory retrieval. *Cerebral Cortex* **8**: 451-461.

Lai, S., Hopkins, A. L., Haacke, E. M., Li, D., Wasserman, B. A., Buckley, P., Friedman, L., Meltzer, H., Hedera, P., and Friedland, R. 1993. Identification of vascular structures as a major source of signal contrast in high resolution 2D and 3D functional activation imaging of the motor cortex at 1.5T: Preliminary results. *Magnetic Resonance in Medicine*, **30**: 387-392.

Lange, N., and Zeger, S. L. 1997. Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging. *Applied Statistician*, **46**: 1-29.

Lawley, D. N. and Maxwell, A. E. 1971. *Factor Analysis as a Statistical Method*. Butterworth and Company, London.

McCullagh, P. and Nelder, J. A. 1983. *Generalized Linear Models* Chapman and Hall, London.

McIntosh, A. R. and Gonzalez-Lima, F. 1994. Structural Equation Modeling and Its Application to Network Analysis in Functional Brain Imaging. *Human Brain Mapping* **2**: 2-22.

McIntosh, A. R., Cabeza, R. E. and Lobaugh, N. J. 1998. Analysis of Neural Interactions Explains the Activation of Occipital Cortex by an Auditory Stimulus. *The American Physiological Society* 2790-2796.

McKeown, M. J., Makeig, S., Brown, G. G., Jung, T.-P., Kindermann, S. S. and Sejnowski, T. J. 1998. Analysis of fMRI by Blind Separation into Independent Spatial Components. *Human Brain Mapping* **6**: 160-188.

Mechelli, A., Penny, W. D., Price, C. J., Gitelman, D. R. and Friston K. J. 2002. Effective Connectivity and Intersubject Variability: Using a Multisubject Network to Test Differences and Commonalities. *NeuroImage* **17**: 1459-1469.

Sunaert, S., Van Hecke, P., Marchal, G., Orban, G. A. 2000. Attention to speed of motion, speed discrimination, and task difficulty: An fMRI study. *NeuroImage* **11**: 612-623.

Warnking, J., Dojat, M., Guerin-Dugue, A., Delon-Martin, C., Olympieff, S., Richard, N., Chehikian, A., Segebarth, C. 2002. fMRI retinotopic mapping - Step by step. *NeuroImage* **17**: 1665-1683.

Zeki, S 1993. *A vision of the brain*. Blackwell, Oxford.

<b>Gamma Measures</b>	<b>Initial 120 individual- voxel datasets</b>	<b>Peak Voxel</b>	<b>Additional 12 datasets for testing the GLM</b>
V1			
Magnitude	1697.43 (1602.35 ; 1792.51)	2462.95	1699.62 (1302.86 ; 2036.38)
Delay	5.16 (5.05 ; 5.28)	5.69	4.94 (4.63 ; 5.25)
Goodness of fit	1.53 (1.49 ; 1.58)	1.44	1.54 (1.38 ; 1.70)
Spread	12.49 (12.01 ; 12.97)	12.41	11.94 (9.17 ; 14.71)
V2			
Magnitude	1912.40 (1749.02 ; 2075.77)	4108.76	1893.18 (1272.75 ; 2513.61)
Delay	4.93 (4.81 ; 5.05)	4.95	4.63 (4.31 ; 4.95)
Goodness of fit	1.60 (1.56 ; 1.65)	1.57	1.56 (1.44 ; 1.68)
Spread	12.34 (11.86 ; 12.81)	13.81	13.45 (11.63 ; 15.26)
V5			
Magnitude	989.06 (915.31 ; 1062.80)	1683.34	1055.04 (904.78 ; 1206.58)
Delay	4.91 (4.83 ; 4.99)	5.68	4.99 (4.78 ; 5.19)
Goodness of fit	1.91 (1.85 ; 1.97)	1.42	1.64 (1.56 ; 1.71)
Spread	11.05 (10.36 ; 11.74)	14.67	14.85 (12.52 ; 17.19)

<b>Covariates</b>	<b><math>\beta</math></b>	<b>P</b>
V1→V2		
V1 Magnitude	0.357	0.009
V2 Magnitude	0.598	< 0.001
V2 Spread	-0.603	< 0.001
V2 Delay <sup>2</sup>	-0.125	0.019
V2 Magnitude <sup>2</sup>	-0.117	0.029
V1 GoF	0.327	0.001
Constant	2.100	< 0.001
V1→V5		
V1 Gof	0.322	0.011
V1 Magnitude*Delay	0.316	0.143
V1 Delay <sup>2</sup>	-0.103	0.124
V5 Magnitude*Delay	0.492	0.130
V2 Delay <sup>2</sup>	0.208	0.014
V5 Delay	0.682	0.056
Constant	0.879	< 0.001
V2→V5		
V2 Delay	-0.208	0.081
V2 Magnitude	0.582	0.011
V2 GoF	0.617	0.001
V1 Magnitude*Delay	-0.467	0.023
V2 Magnitude <sup>2</sup>	-0.143	0.110
V2 Spread	-0.240	0.146
Constant	0.801	< 0.001

## Table Legends

Table 1. Mean and 95% confidence intervals of the measures derived from the gamma curve fits to the data for the initial 120 individual-voxel datasets, for the peak of activation voxel and for the additional 12 datasets for testing the GLM. Delay is measured in seconds, while the other gamma measures have arbitrary units. The values in the brackets represent the 95% confidence interval based on an asymptotic normal distribution. The peak value has no confidence interval since it is a single value.

Table 2. Summary of the final generalised linear models for the three different path connections. The column of  $\beta$  values reports each beta coefficient for the regression, where  $\mathbf{P}$  is the corresponding p-value.

## Figure Legends

Figure 1. Three functional regions selected for modelling of the visual network in the left hemisphere; V1 (red), V2 (blue) and V5 (yellow). Regions are overlaid onto coronal, sagittal and axial views at the co-ordinate  $x = -6$  mm,  $y = -85$  mm,  $z = 12$  mm. The schematic diagram in the lower right panel illustrates the directional connections that were modelled between V1, V2 and V5 for the two levels in the stacked model.

Figure 2. Example of a typical voxel response and its gamma fit. The voxel shown is the peak voxel for V2 ( $T_{645} = 23.72$ ) and it is located at  $x = -12$ ,  $y = -99$ ,  $z = -15.2$  mm. The solid line represents the average time-course calculated by discarding the upper and lower 20% of the values from the drift-corrected data. The dashed line represents the best gamma fit to the data. Also illustrated are the three categorisations of the time-course: stimulus, undershoot and baseline. The first 22 time

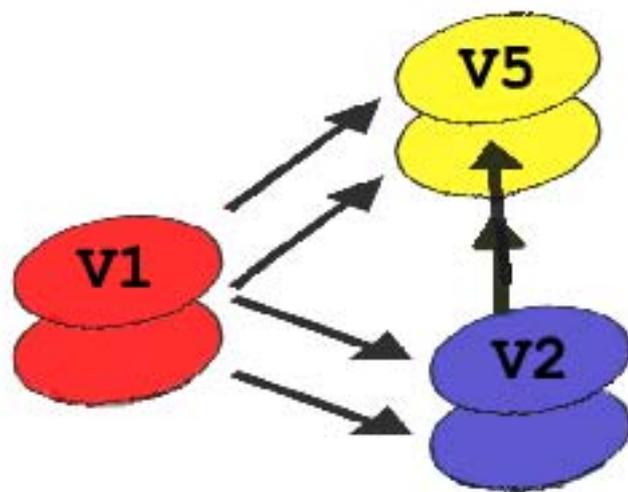
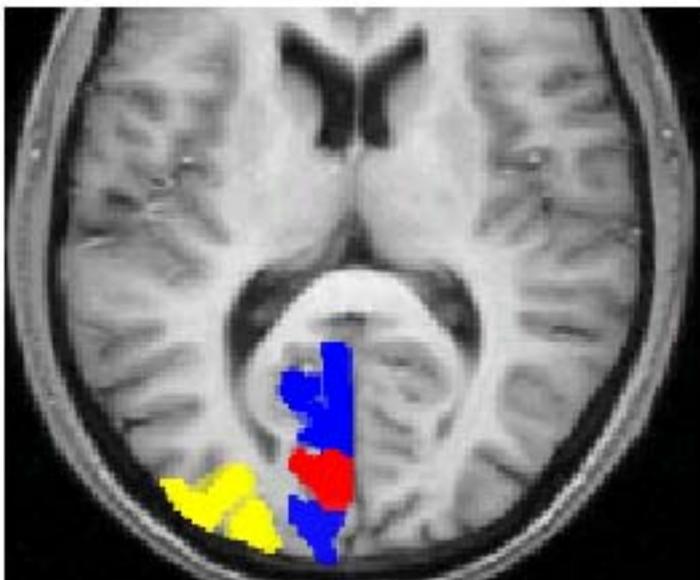
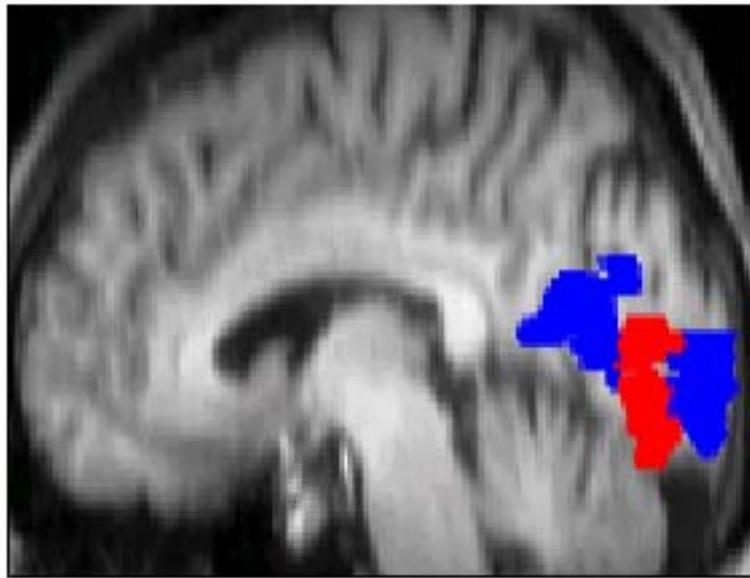
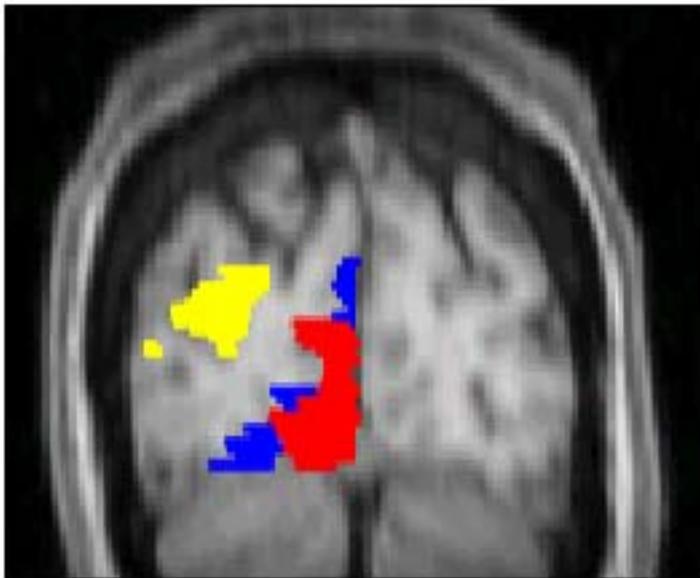
points (0 to 10.5 s) were classified as stimulation, the central 8 time points (11 to 14.5 s) were classified as undershoot and the last 15 time points (15 to and 22 s) were classified as baseline. Also illustrated are the measures of magnitude and delay taken from the gamma fit. The delay is represented by the time that the gamma fit takes to reach its maximum. The magnitude corresponds to the shaded area under the curve.

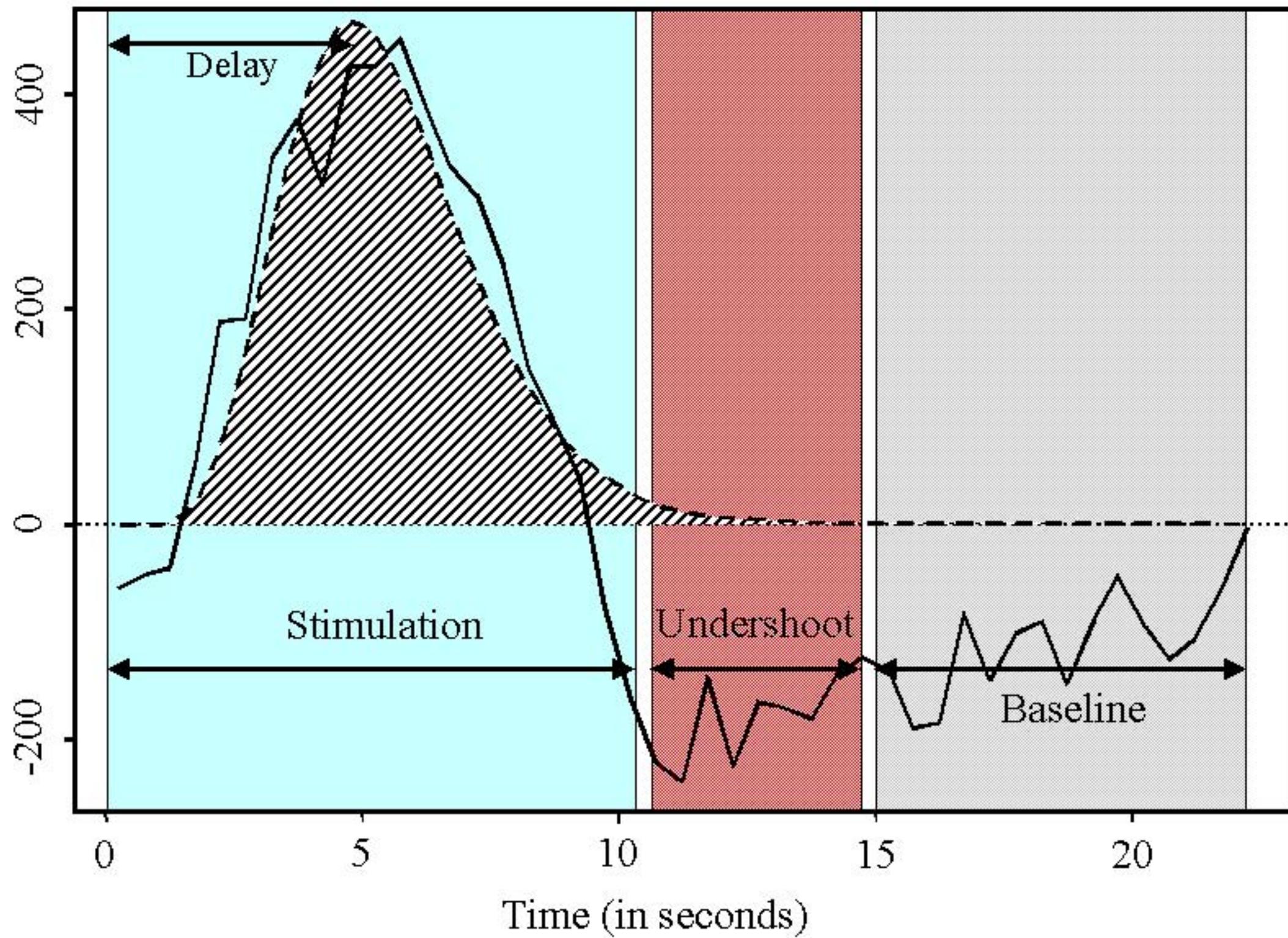
Figure 3. Plot of the coefficients derived from the SEM unconstrained model for the stimulation and baseline conditions for the three connections:  $V1 \rightarrow V2$ ,  $V1 \rightarrow V5$  and  $V2 \rightarrow V5$ . Values above the diagonal line of equality in the diagrams indicate values in which the path coefficient for the stimulation condition is greater than the path coefficient for the baseline condition. Conversely, points below the diagonal of the plot indicate values in which the path coefficient for stimulation condition is weaker than the path coefficient for the baseline condition. The black points represent the values from the individual-voxel dataset where there was no significant condition-specific effects. The red points represent the values from the individual-voxel dataset where there was a significant condition-specific effect. The cyan points represent the connection strengths for stimulation and baseline conditions for the additional 12 datasets used for testing the GLM. The points marked as ‘ $\oplus$ ’ and ‘ $\ominus$ ’ represent the peak voxel and the average of the area within a 5 mm radius of each peak, respectively.

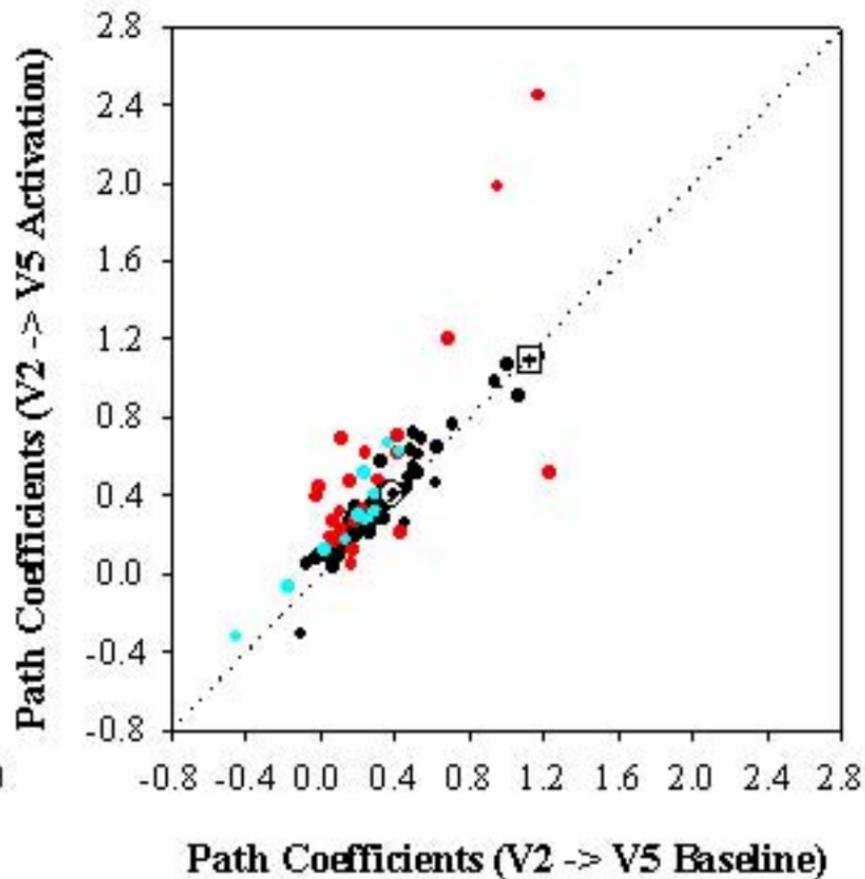
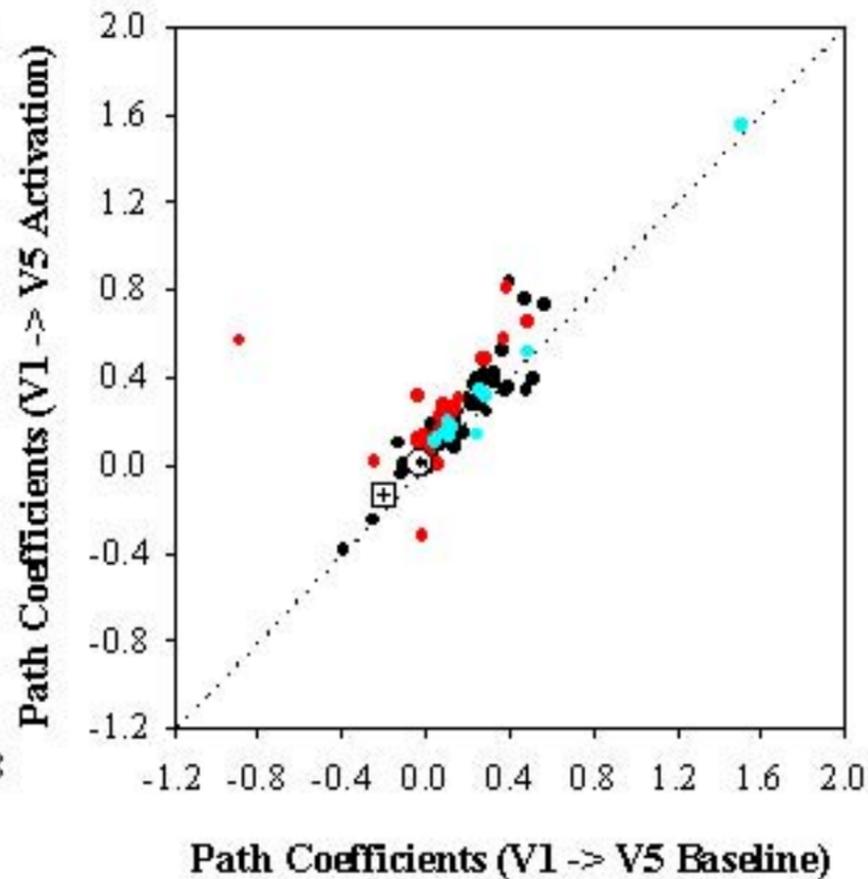
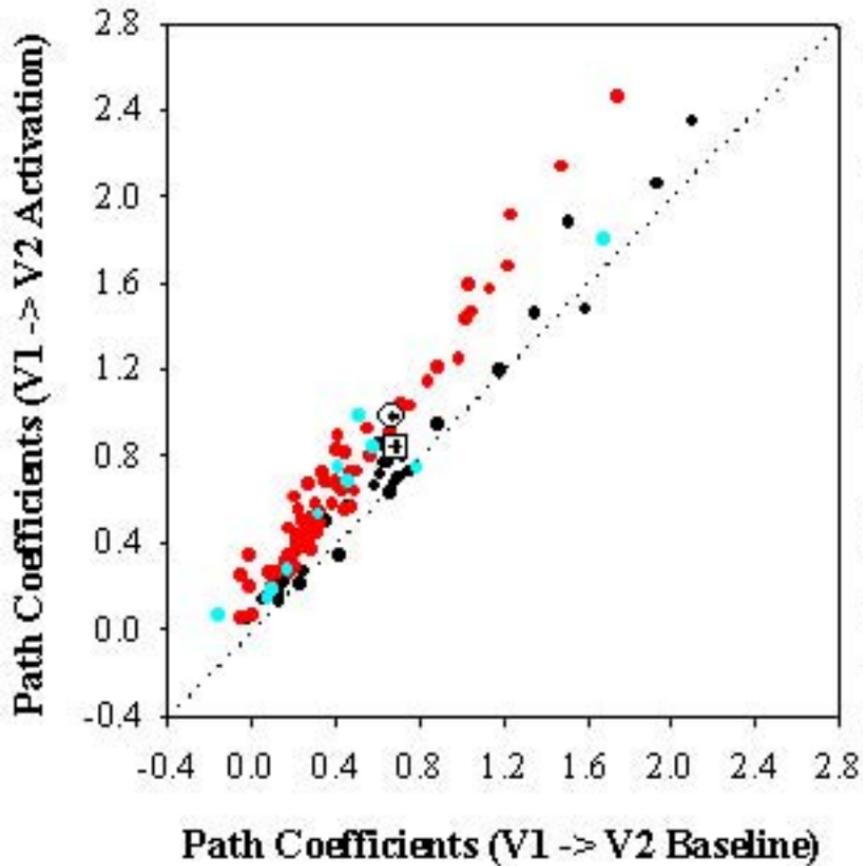
Figure 4. Plot of the chi-square differences derived by subtracting the SEM unconstrained model from SEM constrained model, and the GLM best fit predictions for  $V1 \rightarrow V2$ ,  $V1 \rightarrow V5$  and  $V2 \rightarrow V5$  connections. The black points represent the values from the individual-voxel dataset. The cyan points represent the additional 11 datasets of the 12 datasets used for testing the predictive power of the GLM. One dataset did not converge. The two dotted lines represent the cut-off point of 3.841. All points within the bottom left and the top right quadrants represent a successful fit by the GLM.

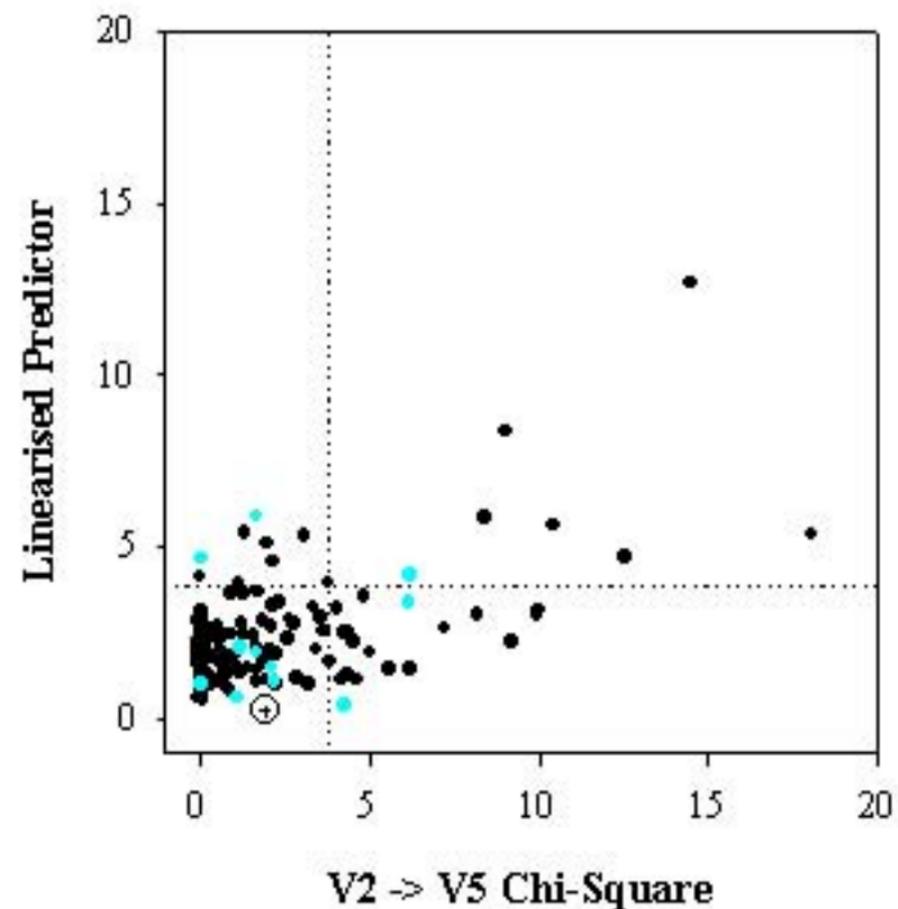
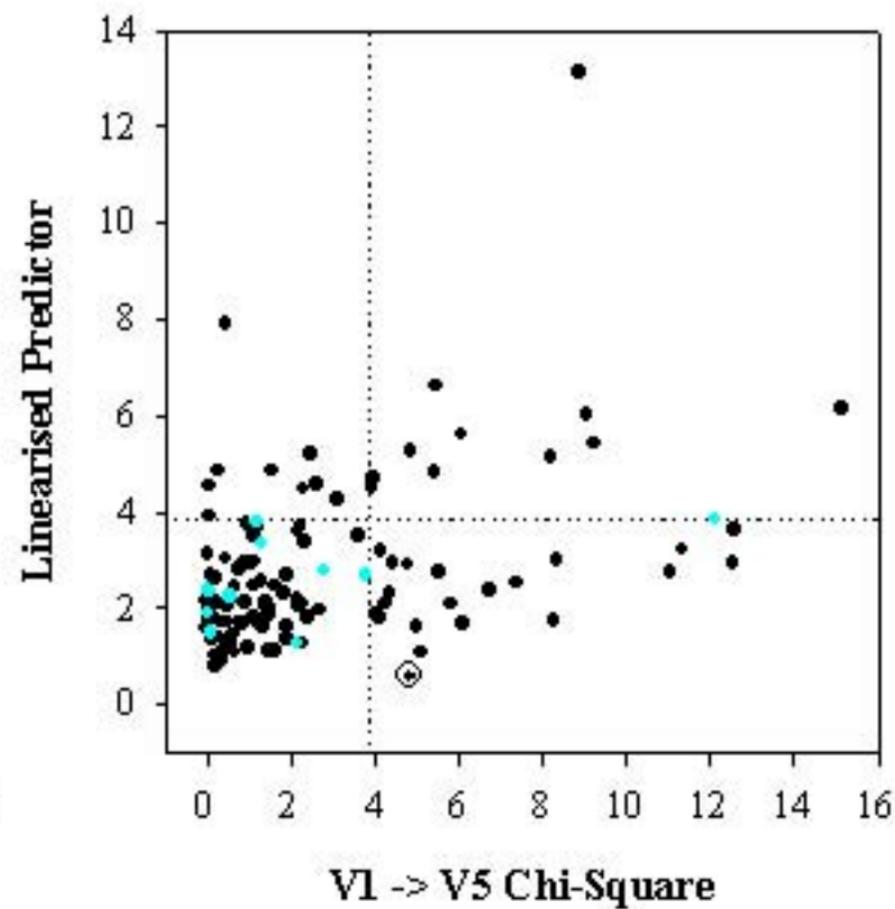
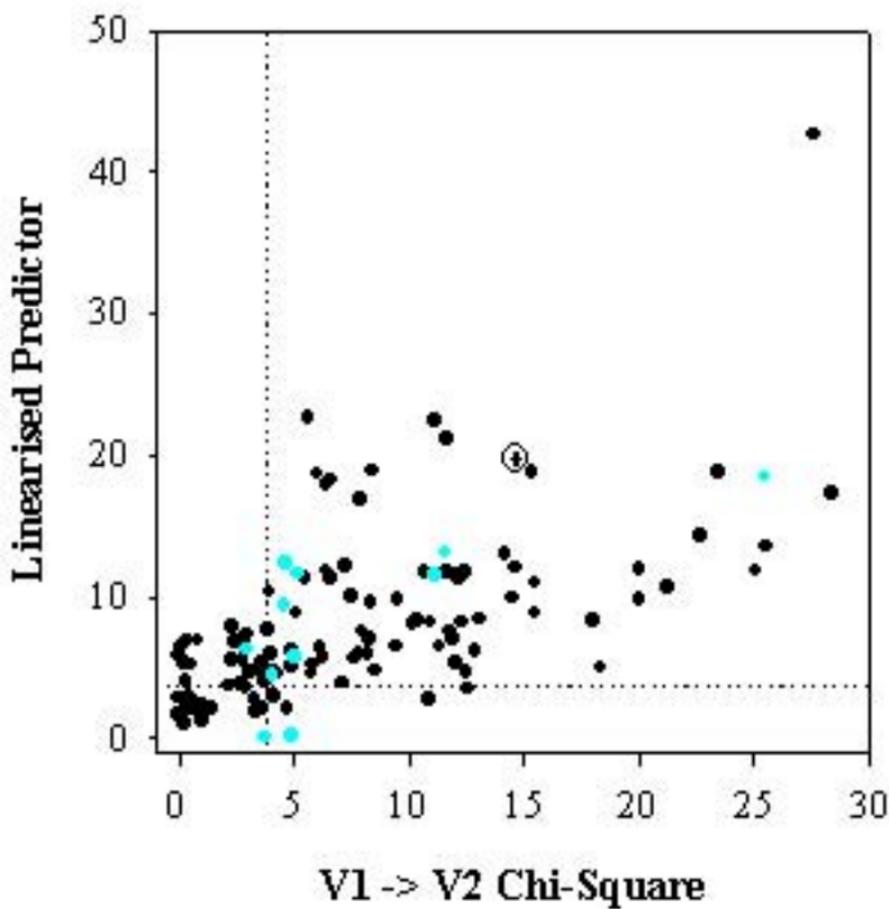
Conversely, all points within the bottom right and the top left quadrants represent a unsuccessful fit by the GLM.

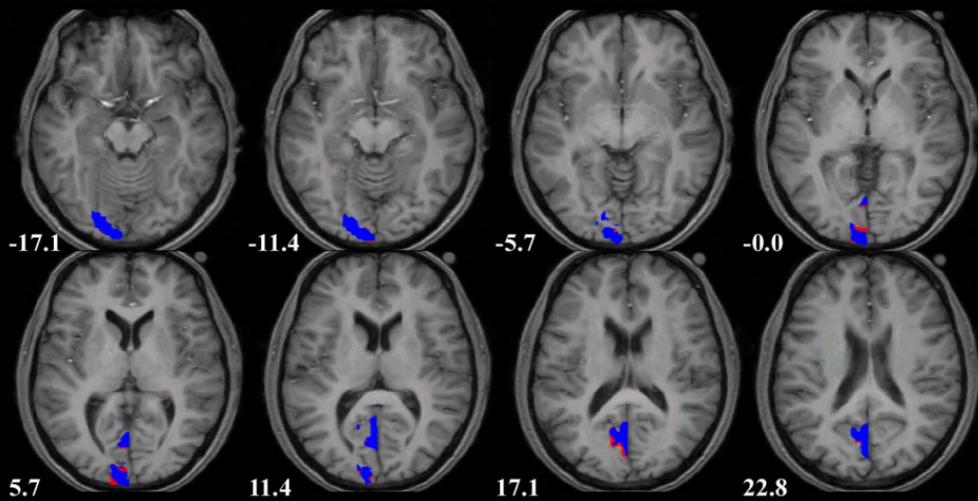
Figure 5. Axial views showing the effects of voxel selection in V2 on the path connection from the SEM modelling for V1→V2 (A) and V2→V5 (B) when the voxels for V1 and V5 are fixed to be the peak of activation. The blue areas indicate voxels in which the difference between the stimulus and baseline conditions were significant ( $\chi^2_{\text{diff}} > 3.841$ ). Conversely, the red areas indicate voxels in which there was no significant condition-specific effects ( $\chi^2_{\text{diff}} < 3.841$ ). The images are displayed for every third brain slice from  $z = -17.1$  mm to  $z = +22.8$  mm; incorporating the location of peak voxel in V2 which is at slice  $z = -15.2$  mm.









**A****B**