



HAL
open science

Dynamic Causal Modelling and physiological confounds: a functional MRI study of vagus nerve stimulation.

Sébastien Reyt, Chloé Picq, Valérie Sinniger, Didier Clarençon, Bruno Bonaz,
Olivier David

► To cite this version:

Sébastien Reyt, Chloé Picq, Valérie Sinniger, Didier Clarençon, Bruno Bonaz, et al.. Dynamic Causal Modelling and physiological confounds: a functional MRI study of vagus nerve stimulation.. *NeuroImage*, 2010, 52 (4), pp.1456-64. 10.1016/j.neuroimage.2010.05.021 . inserm-00498678

HAL Id: inserm-00498678

<https://inserm.hal.science/inserm-00498678v1>

Submitted on 1 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dynamic causal modelling and physiological confounds: A functional MRI study of vagus nerve stimulation

SEBASTIEN REYT^{1,4,+}, CHLOE PICQ^{2,3,4,+}, VALERIE SINNIGER², DIDIER CLARENÇON^{2,3}, BRUNO BONAZ^{2,6}, OLIVIER DAVID^{1,4,5,*}

1. *Neuroimagerie Fonctionnelle et Métabolique, Inserm, U836, Grenoble Institut des Neurosciences, Grenoble, France.*
2. *Stress et Interactions Neuro-Digestives (SIND; EA3744), Inserm, U836, Grenoble Institut des Neurosciences, Grenoble, France.*
3. *Institut de Recherche Biomédicale des Armées, Antenne de La Tronche, Centre de Recherches du Service de Santé des Armées, France.*
4. *Université Joseph Fourier, Grenoble, France.*
5. *Service de Neuroradiologie et Unité IRM, Centre Hospitalier Universitaire, Grenoble, France.*
6. *Service de Gastroentérologie, Centre Hospitalier Universitaire, Grenoble, France.*

⁺ The first two authors contributed equally to this work.

* Corresponding author:

Olivier David
Grenoble Institut des Neurosciences – Inserm U836
Chemin Fortuné Ferrini – Bât EJ Safra
38700 La Tronche, France
Tel: +33 456520586
Fax: +33 456520598
Email odavid@ujf-grenoble.fr

Running Title: DCM of Vagus Nerve Stimulation in fMRI

Revision NIMG-10-406

April 2010

Abstract

Dynamic Causal Modelling (DCM) has been proposed to estimate neuronal connectivity from functional magnetic resonance imaging (fMRI) using a biophysical model that links synaptic activity to hemodynamic processes. However, it is well known that fMRI is sensitive not only to neuronal activity, but also to many other psychophysiological responses which may be task-related, such as changes in cardio-respiratory activity. They are not explicitly taken into account in the generative models of DCM and their effects on estimated neuronal connectivity are not known. The main goal of this study was to report the face validity of DCM in the presence of strong physiological confounds that cannot be corrected for, using an fMRI experiment of vagus nerve stimulation (VNS) performed in rats. First, a simple simulation was used to evaluate the principled ability of DCM to recover directed connectivity in the presence of a confounding factor. Second, we tested the experimental validity using measures of the BOLD correlates of left 5 Hz VNS. Because VNS mostly activates the central autonomic regulation system, fMRI signals were likely to represent both direct and indirect vascular responses to such activation. In addition to the inference of standard statistical parametric maps, DCM was thus used to estimate directed neural connectivity in a small brain network including the nucleus tractus solitarius (NTS) known to receive vagal afferents. Though blood pressure changes may constitute a major physiological confound in this dataset, model comparison of DCMs still allowed the identification of the NTS as the input station of the VNS pathway to the brain. Our study indicates that current developments of DCM are robust to psychophysiological responses to some extent, but does not exclude the need to develop specific models of brain-body interactions within the DCM framework to better estimate neuronal connectivity from fMRI time series.

I Introduction

Dynamic Causal Modelling (DCM) (Friston et al., 2003) has been proposed to estimate neuronal connectivity from functional magnetic resonance imaging (fMRI) time series weighted by the blood oxygen level dependent (BOLD) contrast. It is based on a generative model that links synaptic activity to BOLD time series using a hemodynamic model. Observation equations of the hemodynamic model transform the synaptic activity into a vasodilatory signal, cerebral blood flow, cerebral blood volume and deoxyhaemoglobin content that are combined together to give rise to the BOLD signal (Friston et al., 2000). Synaptic activity amongst different brain regions of interest (ROI) is obtained using state equations that are differential equations derived directly from assumed intrinsic and extrinsic neuronal connectivity (see (Friston et al., 2003) for mathematical details). We have recently shown that estimating this hidden neuronal activity is crucial to identify appropriate orientation of interactions from fMRI time series (David et al., 2008).

DCM has been used in humans to study several cognitive systems such as language (Mechelli et al., 2005), visuospatial attention (Siman-Tov et al., 2007) or face perception (Fairhall and Ishai, 2007). Furthermore, reproducibility of DCM results over scans within the same scan session has been found to be excellent in an experiment on emotional facial expressions (Schuyler et al., 2009). However, it is well known that BOLD fMRI is sensitive not only to task-related neuronal processes, but also to many physiological responses of the central autonomic regulation system (Cechetti and Shoemaker, 2009; Saper, 2002), such as changes in blood pressure related to cardio-respiratory activity (Birn et al., 2006; Dagli et al., 1999; de Munck et al., 2008; Gray et al., 2009; Wise et al., 2004). If they are directly measured, they can be either removed from the data before any connectivity analysis, for instance using subspace projection, or incorporated explicitly in a generative model of the data, such as a DCM, on which connectivity inferences will be performed. When they are task-related however, these physiological responses are correlated to neuronal responses of

interest and thus are difficult to remove during the preprocessing of data. Approaches based on partial correlation, such as partial Granger causality (Guo et al., 2008), then allow to model confounds as latent variables for improved estimation of causal relationships amongst measured time series. Because they are not explicitly taken into account in the generative models of the current release of DCM (www.fil.ion.ucl.ac.uk/spm8, Wellcome Department of Imaging Neuroscience, University College London), the effects of task-related physiological confounds have rarely been considered when interpreting DCM results. The main goal of this study was to report the face validity of DCM for fMRI in the presence of strong physiological confounds that cannot be corrected for, because of lack of experimental measures or lack of differentiability between effects of interest and non interest, using an fMRI experiment of vagus nerve stimulation (VNS) performed in rats.

Although mechanisms of action of VNS remain poorly understood, VNS is used in humans for treatment-resistant depression and epilepsy (Ben Menachem, 2002; Marangell et al., 2007; Milby et al., 2008). It has also been proposed to use VNS as a modulator of intestinal inflammation through the cholinergic anti-inflammatory pathway (Bonaz, 2007; Van Der Zanden et al., 2009). On the one hand, effects observed on the central nervous system are mediated through vagal afferents (80% of VN fibres), which usually are intermittently stimulated at 30 Hz. On the other hand, VN efferent fibres (20%) provide parasympathetic innervation primarily to the lungs, heart, and gastrointestinal tract and have anti-inflammatory properties, *i.e.* the cholinergic anti-inflammatory pathway (Borovikova et al., 2000; Tracey, 2009). Current range frequency of VNS protocols for these purposes is 5-10Hz (Bernik et al., 2002; Naritoku et al., 1995; Tracey, 2007).

Because the nucleus tractus solitarius (NTS), located in the brainstem, receives vagal afferent fibres, central effects of VNS are thought to be mediated through the numerous projections of the NTS into other brain regions. Neurochemical studies (Ben-Menachem et al., 1995; Hamberger et al., 1993; Krahl et al., 1998; Naritoku et al., 1995) suggested activation by VNS of the locus coeruleus and dorsal raphe nucleus, the major sources of

brain norepinephrine and serotonin respectively. This has been confirmed electrophysiologically by showing increases in neuronal firing rates in those structures after acute VNS (Groves et al., 2005) or after long-term stimulation treatment in rats (Dorr and Debonnel, 2006). In humans *in vivo*, central effects of VNS have been studied mostly using electroencephalography (EEG) (Marrosu et al., 2005), positron emission tomography (PET) and single photon emission computer tomography (SPECT) (see (Chae et al., 2003) for a review) and functional magnetic resonance imaging (fMRI) (Bohning et al., 2001; Dietrich et al., 2008; Liu et al., 2003; Lomarev et al., 2002; Nahas et al., 2007; Narayanan et al., 2002; Sucholeiki et al., 2002). Despite a significant variability in reported findings, activations in the medulla/brainstem, limbic regions (insula, anterior cingulate cortex, hippocampus, amygdala, and hypothalamus), thalamus, cerebellum, and periaqueductal grey (PAG) were most commonly observed. In animals, information about VNS mechanisms gathered from neuroimaging studies is scarce (Biraben et al., 2008; Dedeurwaerdere et al., 2005). Because of the modulatory properties of VNS on the central autonomic regulation system, the origins of BOLD effects observed during VNS might not only be glial and neuronal but also be strongly related to other physiological processes. The medullary nuclei targeted by VN indeed couple cardiac function to blood pressure and respiration, the fMRI correlates of which are detectable in humans (Gray et al., 2009). VNS is thus an interesting experimental setup to investigate the behaviour of biophysical models used to analyse fMRI signals to detect neural effect, as proposed in DCM.

Here, we first used a simple simulation to evaluate the principle ability of DCM to recover directed connectivity in the presence of a confounding factor strongly correlated to the experimental design. Second, we tested the experimental validity of this approach using measures of the BOLD correlates of left 5 Hz VNS specifically acquired in Sprague-Dawley rats at 4.7T. In Supplementary Materials, we provide a series of control experiments (heart rate changes during 5 Hz VNS, fMRI of VNS in vagotomised rats, fMRI during ephedrine injection) to ascertain the presence in our dataset of neuronal and non neuronal effects,

through the activation of the central autonomic regulation system. Interestingly, though changes in blood pressure or in heart rate may constitute a major physiological confound in this dataset (see Figure S2 in Supplementary Materials for experimental evidence of tight correlation between BOLD and heart rate), model comparison of families of DCMs still allowed the identification of the NTS as the input station of the VNS pathway to the brain.

II Materials and Methods

II.1 Animals

Adult male Sprague-Dawley rats (250-350g) (Charles River, France) were housed in a controlled environment (12 h light/dark cycles, 20-22°C) with food and water ad libitum. All animals were treated according to guidelines approved by the European Ethics Committee (decree 86/609/CEE). Procedures were approved by the institutional animal welfare committees of the Grenoble Institute of Neuroscience, of the “Institut de Recherche Biomédicale des Armées” and of the legal French Ministry of Agriculture (authorization number 38 04 48).

A total number of 25 rats were used in this study. Eight animals were used for fMRI of VNS. In Supplementary Materials, validation of VNS parameters was done on 6 rats; 4 animals underwent vagotomy and VNS fMRI; 7 animals were used to measure ephedrine effects in fMRI.

II.2 VNS surgery and parameters

Oxygen saturation and heart rate were monitored during surgery. Anaesthesia was induced with 5% isoflurane in air/oxygen mixture (8:2) and maintained during surgery at 2.5%. An incision was made over the ventral region of the neck. Then the left (or right for the first experiment comparing left and right VNS) VN and the carotid artery were isolated. A bipolar electrode (Cyberonics Inc., Webster, Texas, USA) was gently wrapped around the pack and

connected to a stimulation chain: S88 stimulator, SIU5 isolation unit, CCU1 constant current unit (Grass Technologies, Astro-Med, RI, US). VNS was performed using biphasic pulses with the following parameters: 1mA, 5Hz, 500 μ s pulse width and cycle of 9 s ON / 51 s OFF.

II.3 Animal preparation for fMRI

Immediately after electrode implantation, a bolus of 0.5 mL/kg dexmedetomidine (Dexdomitor, Pfizer) was administered in the hind limb muscle on a 1 minute basis, and isoflurane was discontinued to maximize the specificity of BOLD activations (Weber et al., 2006). Animals were then placed in a rat holder (Bruker, Germany) and fixed with tooth bar and ear plugs to prevent any motion. Respiratory rate, temperature and heart rate were visually monitored throughout fMRI experiments. ECG was recorded using two gold plated electrodes and conductive gel fixed with micropore surgical tape placed on the right forelimb and the left hind limb. Experiments where animals presented abnormal respiratory patterns were discontinued and these animals were removed from the study. Body temperature was maintained by a water-circulating heating-pad. All physiological measures were done with a MR-compatible small animal monitoring and gating system (Model 1025, SA Instruments Inc., Stony Brook, NY, USA). The electrode was connected to the stimulator, and animals were inserted in the magnet. Animals were freely breathing an air/O₂ mixture (9:1) during fMRI experiments that lasted no longer than 60 minutes. At the end of fMRI experiments, animals were anesthetized with 2.5% isoflurane in air/O₂ mixture during removal of electrode. They were then euthanized with an intracardiac injection of Pentobarbital (Dolethal®; 200 mg/kg).

II.4 MR parameters

fMRI experiments were performed on a 4.7 T Bruker Avance III horizontal animal scanner, equipped with a 21 cm shielded gradient system (600mT/m). Radiofrequency transmission was achieved with a 12 cm diameter volume coil, and the resonance signal was received

with a surface coil placed over the head of the animal to allow an optimal recording of the brain and brainstem. Both coils were actively decoupled from each other.

FLASH pilot slices in three directions were acquired to position the animal in the magnet with the plane of the interaural line at the magnet center. A field map was then acquired, and shimming on whole brain was performed using MapShim (Bruker). Final shim was within 50 Hz on the entire brain for each animal. Anatomical images were acquired for brain normalization using a RARE sequence: TE = 13.125 ms; effective 52.5 ms; TR = 5800 ms; RARE factor = 8; acquisition time = 3min30s.

Once all physiological parameters were stabilized, three fMRI sessions of 10 min duration each were acquired. BOLD signals were obtained using coronal multi-slice gradient-echo EPI with the following parameters: TE = 30 ms (echo position 33%, no zero-filling); TR = 3 s; BW = 200 kHz; flip angle 75; 25 consecutive slices of 1 mm thickness; FOV = 25.6x25.6 mm; matrix of 64x64 pixels, allowing a spatial resolution of 0.4x0.4x1 mm.

A standard block design was used for VNS, alternating ON and OFF periods expressed in scans as follows: 10 OFF + 9 x [3 ON + 17 OFF] + 3 ON + 7 OFF. Duration of each session was 10 min. Animals were allowed to rest for at least 7 minutes between two successive sessions.

II.5 FMRI data analysis

Standard fMRI data analysis was done using SPM5 (www.fil.ion.ucl.ac.uk/spm, Statistical Parametric Mapping, Wellcome Department of Imaging Neuroscience, Functional Imaging Laboratory, London, UK). Some routines of this software were adapted to rat imaging in accordance with (Schweinhart et al., 2003). DCM analyses were performed using SPM8.

II.5.1 Preprocessing

For each session, EPI volumes were first realigned to account for motion correction. Then all images were normalised to a template with coordinates chosen according to the rat atlas of Paxinos and Watson, with the origin at the bregma (Paxinos and Watson, 1997). Practically,

we used the T2 template kindly provided by Schweinhardt et al., who have provided useful guidelines to construct specific small animal templates on any MR system (Schweinhardt et al., 2003). Normalised images were resampled to reach an isotropic spatial resolution of 0.4 mm. Finally, normalised EPI images were smoothed with a Gaussian kernel of 0.8 mm FWHM, and detrended using a high-pass filter based on the discrete cosine transform (standard cutoff of 1/128s).

II.5.2 Statistical maps of BOLD changes

Statistical analysis of BOLD signals was done on smoothed, normalised and realigned EPI images. The regressor of interest was obtained by convolving the VNS paradigm with the canonical hemodynamic response function provided in SPM5. For each animal, maps of the t-statistic of activations were obtained by correlating the high-pass filtered time series of each voxel with the regressor of interest using a standard first-level statistical design (Friston et al., 2005). Activations at the group level were obtained using a fixed-effect analysis following guidelines provided in (Friston et al., 1999). The decision to perform a fixed-effect analysis was based (i) on the reduced number of animals in each condition, which was too small to perform a random-effect analysis, and (ii) on the good reproducibility between animals of the activation patterns.

II.5.3 Dynamic Causal Modelling

In DCM for fMRI (Friston et al., 2003), a bilinear neuronal state equation specifies the connectivity between brain regions. The synaptic activity is then transformed into BOLD signals using a hemodynamic model. Knowing the data and the stimulus input (VNS paradigm), DCM proceeds to a conjoint estimation, from the measured BOLD time series, of the neuronal connectivity and of the hemodynamic parameters. The evidence of a dynamic causal model is defined by its marginal likelihood obtained after model optimisation (Penny et al., 2004). Comparing competing models can thus readily be performed by comparing the model evidence (the most plausible model is the one with the largest evidence). Here, the

model log-evidence was approximated by the model negative free energy. The free energy is the quantity (objective function) optimised by variational Bayes methods implemented in SPM8 for DCM (Friston et al., 2007). It provides a bound on the log-evidence for any model (exact for linear models). Assuming each data set is independent of the others and that only one model underlies brain responses, the log-evidence at the group level in a fixed effect (FFX) analysis is simply obtained by adding the log-evidence of each session (Garrido et al., 2007). To diminish the sensitivity to outliers, which may be important in small group samples, and to the possibility that various brains react differently to the same stimulus, a methodology for random effect (RFX) analysis of group model evidences has recently been proposed (Stephan et al., 2009). We performed both FFX and RFX analyses on actual data for comparing group evidence using DCM code as implemented in SPM8.

When many models are compared together, finding “the best model” may be inadequate because of the large number of experimental factors and modelling approximations that remain insufficiently controlled to detect small differences amongst models. To overcome this shortcoming, it has been proposed very recently to combine family level inference and Bayesian model averaging within families (Penny et al., In press). Family level inference indeed removes uncertainty about aspects of model structure other than the characteristic of interest. Here, we tested whether DCM was able to recover two aspects from known neuroanatomy: (i) the NTS receives vagal afferents; (ii) interregional connectivity is largely bidirectional.

To estimate whether VNS activations could be explained by propagation of NTS neural activity, after receiving all vagal afferents, the time series of deactivated regions by VNS were first extracted. Second, according to observed deactivations, five ROIs were delineated from the rat atlas of Paxinos and Watson (Paxinos and Watson, 1997) and were chosen as follows (Figure 1 & Table 1): NTS, cingulate cortex (Cg), retrosplenial granular cortex (RSG), anterior hippocampus (restricted to CA1), and insula (AI). ROI time series were obtained for each session using the SPM ROI tool, which computes the first scaled

eigenvariate (i.e. weighted mean) of the time series of voxels (between 130 and 200 voxels per ROI, on average) selected by applying the anatomical mask of the ROI (Figure 1) on the map of negative t-values, thresholded at $p < 0.001$ uncorrected. Finally, 410 DCMs were generated as follows, in a constrained way because it was not practically feasible to test all possible models: Given 5 ROIs, there exist 41 possible graphs which connect unidirectionally ROI 1 to the others (with serial or parallel connections, ROI 1 receiving an extrinsic input modelling VNS afferents). Assuming that all regions could potentially receive the VNS input, $41 \times 5 = 205$ unidirectional models with a single extrinsic input were generated and estimated. The location of the extrinsic input defined a family of model (named according to the input station) and the point (i) above (whether DCM for fMRI estimated the NTS as being the input station of VNS) was addressed by using Bayesian model comparison (FFX and RFX) of the five families. In addition, 205 bidirectional models were generated from the previous 205 unidirectional models and **fitted** to the data. By comparing the two families defined by unidirectional and bidirectional connections, the point (ii) above (whether DCM estimates bidirectional long range brain connectivity) was finally investigated.

Figure 1 & Table 1 about here

II.6 Simulation

To evaluate the sensitivity of DCM to detect causal relationships as a function of the amplitude of confounds, we set up a toy model: 2 ROIs were considered (1 & 2) that were mutually connected with a weight of 0.7. Time series of 10 min duration were generated with parameters similar to those in Experimental Section II.4: 180 scans, TR = 3 s, input u on ROI 1 same as the VNS function. Time series y_1 and y_2 of ROIs 1 & 2 were generated by integrating the following state equations:

$$\begin{cases} \dot{x}(t) = \begin{bmatrix} -1 & 0.7 \\ 0.7 & -1 \end{bmatrix} x(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t) \\ y(t) = h(x(t)) + \sigma \varepsilon(t) \end{cases} \quad (1)$$

where $\dot{x} = \frac{dx}{dt}$, t is time, $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ are the neuronal states and $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ are BOLD time series obtained from x using the hemodynamic model h with canonical parameters as described in (Stephan et al 2007). Normally distributed random noise with unit variance, $\varepsilon(t)$, was assumed to model physiological and experimental noise uncorrelated to stimulation $u(t)$. Positive parameter σ allowed to tune signal to noise ratio (SNR).

Physiological confounds c were defined as:

$$c = \frac{y_1 + y_2}{2}. \quad (2)$$

They were thus highly correlated to BOLD time series y generated by neuronal activity x .

Measured time series m were assumed to be a linear combination of y and c :

$$m = y + \begin{bmatrix} r \\ r \end{bmatrix} c \quad (3)$$

where r is the confound to neuronal hemodynamic response ratio. It was then varied from 0 (no confound) to 10. SNR was defined as the ratio between the standard deviations of $m(t)$ and of $\sigma \varepsilon(t)$, and was varied between 5 and 1. For each tested values of the pair $[r, SNR]$, generated time series $m(t)$ were fitted using two models: the true model (model A) as described by the neuronal state equation (1), and two false models in which the extrinsic input impinged on ROI 2 only (model B), or on both ROIs (model C) (Figure 2).

The influence of the confounds c on DCM ability to detect true directionality of information flow between ROI 1 and ROI 2 was evaluated by plotting the difference in log model evidence (approximated by the negative free energy) of models A, B and C, as a function of r and of SNR.

III Results

III.1 Simulation

Figure 2A shows two minutes of simulated time series with hemodynamic response function (HRF) amplitude normalised to 1 for display. The “true” generative model used here is defined by Equation (1). The three first panels show different degrees of contamination by confounds (common mode of original time series) defined by different values of parameter r , without random noise ($SNR = Inf$, *i.e.* $\sigma = 0$). As one might expect from the design of this simple simulation, increasing r induced a reduction of the time delay between the activity of ROIs 1 & 2, and therefore challenged the detection of information flow from ROI 1 towards ROI 2. The three last panels show examples of simulated BOLD time series for the different values of SNR.

Figure 2B shows the difference of negative free energy (approximate of log model evidence) between Models A & B (top) and Models A & C (bottom), when fitted to data shown in Figure 2A and averaged over 50 realisations, as a function of parameter r and SNR . When r increased, model evidences decreased (global effect of confounds not shown) and also the difference of log evidence between the Models A & B. The effect of random noise (quantified by the SNR) was similar to that of correlated confound. Model A (true model) always remained more plausible than Model B (indicated a false position of the extrinsic input), as indicated by its largest log evidence. However, the difference between Models A & C was not significant (a value of 3 if log evidence being a common threshold value for significance), whatever the degree of noise considered.

Figure 2 about here

III.2 fMRI of VNS

Highly significant VNS-related deactivations ($p < 0.001$, FWE corrected) were found at the animal level and at the group level ($n=8$) (Figure 3). No significant activations were observed, neither at the group level nor at the animal level. While most of the brain was found to be deactivated at the group level, the strongest deactivations were mainly located in subcortical structures, such as the anterior NTS, the locus coeruleus, the parabrachial nucleus and the vestibular nucleus, in the anterior hippocampus, in the inferior and superior colliculi, or in the cerebellum. Most significant cortical deactivations were found in the retrosplenial, prelimbic, infralimbic, anterior cingulate and agranular insular cortices.

Figure 3 about here

Three rats were removed from DCM analyses because the signal-to-noise ratio was too poor in the predefined ROIs. In the remaining five rats, BOLD time series recorded during stimulation of the intact left VN were well fitted by the different DCMs of Figure 1. Bayesian model selection was performed at the group level on families of models (Penny et al., In press) to test two main features (Figure 4). First (Figure 4 left), we addressed the identification of the input station of VNS by grouping together the 41 models having the same extrinsic input. We found a clear preference for models receiving vagal afferents on the NTS (FFX posterior probability: AI=0, CA1=0, Cg=0, NTS=1, RSG=0; RFX exceedance probability: AI=0.0003, CA1=0.0002, Cg=0.1474, NTS=0.8493, RSG=0.0028). Second (Figure 4 right), we verified that the DCM approach was capable to detect the presence of bilateral long range connectivity by comparing the 210 models with bilateral connections against the 210 other models with forward connections only (FFX posterior probability: Unidirectional=0, Bidirectional=1; RFX exceedance probability: Unidirectional=0, Bidirectional=1). Finally, the most plausible model (Figure 4 middle) was identified as the one

combining those two features, with the fastest spreading from the NTS towards other regions (i.e. direct connections).

Figure 4 about here

IV Discussion

To our knowledge, the central effects of VNS at a low frequency of stimulation have been rarely explored (Lomarev et al., 2002; Osharina et al., 2006). Here we provided the first fMRI study of acute VNS performed in rodents. Highly significant VNS-related deactivations were found in large portions of the brain, and particularly in the NTS and closely connected structures, such as the parabrachial nucleus, the locus coeruleus and the hippocampus. The VNS-induced fMRI deactivation of the cerebellum was also in agreement with the known anatomical projections of the NTS to the inferior portions of the cerebellum (Xu and Frazier, 1995). Significant deactivations were also reported in the prefrontal cortex and retrosplenial cortex, regions which are known to express c-Fos protein, an indirect marker of basal level of neuronal firing, after continuous 30 Hz VNS (Naritoku et al., 1995).

The observed massive deactivation of the brain is different from reported findings in the only other study of VNS using functional neuroimaging published in rat (Dedeurwaerdere et al., 2005), where glucose metabolism, as measured by FDG-PET, significantly changed during acute VNS only in the hippocampus (decrease) and olfactory bulbs (increase). In patient studies of acute VNS, local activations and deactivations were also reported (Bohning et al., 2001; Dietrich et al., 2008; Liu et al., 2003; Lomarev et al., 2002; Nahas et al., 2007; Narayanan et al., 2002; Sucholeiki et al., 2002). The differences between these findings and ours can be explained by several factors, such as diversity in stimulation protocols or in animal preparation. It is also difficult to directly compare results obtained in rats and in

humans. In fact, two elements led us to consider a strong cardio-respiratory contribution to these deactivations: (i) deactivation of the brain was concomitant with a transient decrease of heart rate (6% of heart rate decrease measured during 1 h left VNS, see Supplementary Materials); (ii) the most significantly deactivated structures belonged to the central autonomic regulation system (Saper, 2004). Two experimental controls to further understand BOLD statistical maps of acute VNS were performed (see Supplementary Materials): (i) No brain activation was found when vagal afferents were cut, whereas brain activations remained unchanged when vagal efferents were disrupted. Critically, this series of oriented vagotomy experiments demonstrated a major role taken by afferent fibres in comparison to efferent fibres, and thus suggested the necessity of the activation of regulation nuclei of the mesencephalon to explain BOLD responses to VNS. (ii) BOLD responses to ephedrine injection showed a significant positive correlation with heart rate changes (ephedrine increases heart rate (Avois et al., 2006)) in many brain regions also deactivated during VNS. VNS and ephedrine results put together, they suggested that the negative sign of BOLD activations during VNS was mainly driven by heart rate decrease.

Because weak bradycardia accompanied VNS, it remained difficult to distinguish neuronal effects and cardio-respiratory effects, which are known to participate significantly to BOLD signals (Birn et al., 2006; Dagli et al., 1999; de Munck et al., 2008; Gray et al., 2009; Wise et al., 2004). DCM as used in this study can be perceived as a way to investigate whether there were neuronal components into measured fMRI signals. Indeed if inferred neuronal connectivity agreed with known neuronal circuitry, *i.e.* the NTS receives afferent inputs, then an indirect demonstration of the neural nature of measured BOLD signals would have been provided. Model comparison at the group level showed it was the case since the NTS region was the most likely structure amongst five candidates to receive VNS inputs, in agreement with strong knowledge about neuroanatomical underpinnings of VNS. Moreover, DCM results were reinforced by the fact that we found clear evidence for bilateral long range connectivity, as commonly found in neuroanatomical studies. The best model at the group level (Figure 4) indicated fast (*i.e.* direct) connections from the NTS towards other brain regions. This is very reminiscent of the view of the NTS as a “hub”, or input station to the

brain. It also indicates that DCM finds effective connectivity between regions that are not anatomically directly connected. This may be the case for CA1, which is known to receive signals from the NTS, but probably through the amygdala or the locus coeruleus (which were not included in the DCMs and therefore acted as latent variables here).

Along with other electrophysiological validation studies of DCM combined with fMRI (David et al., 2008) or microdialysis recordings (Moran et al., 2008), this new experimental validation of DCM for fMRI is important for two reasons: (i) it is the second demonstration of the ability of DCM for fMRI to estimate neuronal connectivity features in agreement with prior knowledge in rat datasets and thus confirms conclusions derived in (David et al., 2008); (ii) it shows that DCM is able to recover the presumed neural input region in the presence of physiological noise. This may be because the state equations of the hemodynamic model are well suited to fit smooth physiological responses and that priors on hemodynamic parameters are sufficiently relaxed to allow to do so, *i.e.* to capture neural or indirect hemodynamic variability (David et al., 2008; Friston, 2009; Roebroeck et al., 2009).

To further illustrate this point, we performed a short simulation on a toy model composed of two interconnected regions where a hemodynamic confound modelling task-related physiological responses was assumed to be symmetrically correlated to the neurally induced hemodynamic activity of each region (Figure 2). The evidence of DCMs was obviously sensitive to the presence of correlated confounds. However, it was possible to detect the true model as being the most plausible model having generating the data when the models used distinct extrinsic inputs (Models A & B). The limit of significance was nonetheless reached for strong correlated confounds ($r > 300\%$), for different degrees of uncorrelated random noise (experimental or physiological). When the models shared common inputs (and probably extrinsic inputs that bypass true intrinsic connectivity such as in Model C), DCM was not able to differentiate models from average log evidence, for any noise level. This interesting result is reassuring for DCM analyses in fMRI, in general, as long as the location of extrinsic inputs is well motivated on the basis of strong anatomical

knowledge of brain afferents. However, in other configurations, such as asymmetric confounds, additive experimental noise or more complex DCM architectures with potential interactions between intrinsic (*i.e.* between brain regions) and extrinsic (*e.g.* stimuli) connectivity, this ideal scenario could potentially be challenged and reversed.

Robustness of DCM in the presence of physiological responses as demonstrated here may not be true for all studies. Indeed, psychophysiological measurements, such as cardiac, vascular, respiratory, electrodermal, gastrointestinal, and papillary monitoring, may reveal subtle processes during brain-body interactions that are modulated by the experimental context, *e.g.* emotion or arousal level (Gray et al., 2009). Because DCM aims at revealing changes of neural connectivity associated to differential contextual responses, changes in physiological responses may be wrongly interpreted as changes in neuronal weights in the absence of an adapted generative model of brain-body interactions. Developing such models may be very important for future fMRI studies completed by parallel physiological monitoring to ascertain the validity of a neuronal interpretation of BOLD activation maps. This task is exciting but very complex and was well beyond the scope of this paper. Hopefully, fMRI modellers would soon come up with a solution to this problem.

V Acknowledgments

We thank Cyberonics France for providing us with the VNS electrodes. This study was funded by Inserm and Agence Nationale pour la Recherche. The authors declare no conflict of interest.

References

- Avois, L., Robinson, N., Saudan, C., Baume, N., Mangin, P., Saugy, M., 2006. Central nervous system stimulants and sport practice. *Br J Sports Med* 40 Suppl 1, i16-20.
- Ben Menachem, E., 2002. Vagus-nerve stimulation for the treatment of epilepsy. *Lancet Neurol.* 1, 477-482.
- Ben-Menachem, E., Hamberger, A., Hedner, T., Hammond, E.J., Uthman, B.M., Slater, J., Treig, T., Stefan, H., Ramsay, R.E., Wernicke, J.F., et al., 1995. Effects of vagus nerve stimulation on amino acids and other metabolites in the CSF of patients with partial seizures. *Epilepsy Res* 20, 221-227.
- Bernik, T.R., Friedman, S.G., Ochani, M., DiRaimo, R., Susarla, S., Czura, C.J., Tracey, K.J., 2002. Cholinergic antiinflammatory pathway inhibition of tumor necrosis factor during ischemia reperfusion. *J Vasc Surg* 36, 1231-1236.
- Biraben, A., Guerin, S., Bobillier, E., Val-Laillet, D., Malbert, C.-H., 2008. Central activation after chronic vagus nerve stimulation in pigs: Contribution of functional imaging. *Bull Acad Vét France* 161, 441-448.
- Birn, R.M., Diamond, J.B., Smith, M.A., Bandettini, P.A., 2006. Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. *Neuroimage* 31, 1536-1548.
- Bohning, D.E., Lomarev, M.P., Denslow, S., Nahas, Z., Shastri, A., George, M.S., 2001. Feasibility of vagus nerve stimulation-synchronized blood oxygenation level-dependent functional MRI. *Invest Radiol* 36, 470-479.
- Bonaz, B., 2007. The cholinergic anti-inflammatory pathway and the gastrointestinal tract. *Gastroenterology* 133, 1370-1373.
- Borovikova, L.V., Ivanova, S., Zhang, M., Yang, H., Botchkina, G.I., Watkins, L.R., Wang, H., Abumrad, N., Eaton, J.W., Tracey, K.J., 2000. Vagus nerve stimulation attenuates the systemic inflammatory response to endotoxin. *Nature* 405, 458-462.
- Cechetto, D.F., Shoemaker, J.K., 2009. Functional neuroanatomy of autonomic regulation. *Neuroimage* 47, 795-803.
- Chae, J.H., Nahas, Z., Lomarev, M., Denslow, S., Lorberbaum, J.P., Bohning, D.E., George, M.S., 2003. A review of functional neuroimaging studies of vagus nerve stimulation (VNS). *J Psychiatr Res* 37, 443-455.
- Dagli, M.S., Ingeholm, J.E., Haxby, J.V., 1999. Localization of cardiac-induced signal change in fMRI. *Neuroimage* 9, 407-415.

- David, O., Guillemain, I., Saillet, S., Reyt, S., Deransart, C., Segebarth, C., Depaulis, A., 2008. Identifying neural drivers with functional MRI: an electrophysiological validation. *PLoS Biol* 6, 2683-2697.
- de Munck, J.C., Goncalves, S.I., Faes, T.J., Kuijjer, J.P., Pouwels, P.J., Heethaar, R.M., Lopes da Silva, F.H., 2008. A study of the brain's resting state based on alpha band power, heart rate and fMRI. *Neuroimage* 42, 112-121.
- Dedeurwaerdere, S., Cornelissen, B., Van Laere, K., Vonck, K., Achten, E., Slegers, G., Boon, P., 2005. Small animal positron emission tomography during vagus nerve stimulation in rats: a pilot study. *Epilepsy Res* 67, 133-141.
- Dietrich, S., Smith, J., Scherzinger, C., Hofmann-Preiss, K., Freitag, T., Eisenkolb, A., Ringler, R., 2008. [A novel transcutaneous vagus nerve stimulation leads to brainstem and cerebral activations measured by functional MRI]. *Biomed Tech (Berl)* 53, 104-111.
- Dorr, A.E., Debonnel, G., 2006. Effect of vagus nerve stimulation on serotonergic and noradrenergic transmission. *J Pharmacol Exp Ther* 318, 890-898.
- Fairhall, S.L., Ishai, A., 2007. Effective connectivity within the distributed cortical network for face perception. *Cereb Cortex* 17, 2400-2406.
- Friston, K., 2009. Dynamic causal modeling and Granger causality Comments on: The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. *Neuroimage*.
- Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2007. Variational free energy and the Laplace approximation. *Neuroimage* 34, 220-234.
- Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. *Neuroimage* 19, 1273-1302.
- Friston, K.J., Holmes, A.P., Worsley, K.J., 1999. How many subjects constitute a study? *Neuroimage* 10, 1-5.
- Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. *Neuroimage* 12, 466-477.
- Friston, K.J., Stephan, K.E., Lund, T.E., Morcom, A., Kiebel, S., 2005. Mixed-effects and fMRI studies. *Neuroimage* 24, 244-252.
- Garrido, M.I., Kilner, J.M., Kiebel, S.J., Stephan, K.E., Friston, K.J., 2007. Dynamic causal modelling of evoked potentials: a reproducibility study. *Neuroimage* 36, 571-580.
- Gray, M.A., Minati, L., Harrison, N.A., Gianaros, P.J., Napadow, V., Critchley, H.D., 2009. Physiological recordings: basic concepts and implementation during functional magnetic resonance imaging. *Neuroimage* 47, 1105-1115.
- Groves, D.A., Bowman, E.M., Brown, V.J., 2005. Recordings from the rat locus coeruleus during acute vagal nerve stimulation in the anaesthetised rat. *Neurosci Lett* 379, 174-179.

- Guo, S., Seth, A.K., Kendrick, K.M., Zhou, C., Feng, J., 2008. Partial Granger causality--eliminating exogenous inputs and latent variables. *J Neurosci Methods* 172, 79-93.
- Hamberger, A., Haglid, K., Nystrom, B., Silfvenius, H., 1993. Co-variation of free amino acids in human epileptogenic cortex. *Neurochem Res* 18, 519-525.
- Krahl, S.E., Clark, K.B., Smith, D.C., Browning, R.A., 1998. Locus coeruleus lesions suppress the seizure-attenuating effects of vagus nerve stimulation. *Epilepsia* 39, 709-714.
- Liu, W.C., Mosier, K., Kalnin, A.J., Marks, D., 2003. BOLD fMRI activation induced by vagus nerve stimulation in seizure patients. *J Neurol Neurosurg Psychiatry* 74, 811-813.
- Lomarev, M., Denslow, S., Nahas, Z., Chae, J.H., George, M.S., Bohning, D.E., 2002. Vagus nerve stimulation (VNS) synchronized BOLD fMRI suggests that VNS in depressed adults has frequency/dose dependent effects. *J Psychiatr Res* 36, 219-227.
- Marangell, L.B., Martinez, M., Jurdi, R.A., Zboyan, H., 2007. Neurostimulation therapies in depression: a review of new modalities. *Acta Psychiatr Scand* 116, 174-181.
- Marrosu, F., Santoni, F., Puligheddu, M., Barberini, L., Maleci, A., Ennas, F., Mascia, M., Zanetti, G., Tuveri, A., Biggio, G., 2005. Increase in 20-50 Hz (gamma frequencies) power spectrum and synchronization after chronic vagal nerve stimulation. *Clin Neurophysiol* 116, 2026-2036.
- Mechelli, A., Crinion, J.T., Long, S., Friston, K.J., Lambon Ralph, M.A., Patterson, K., McClelland, J.L., Price, C.J., 2005. Dissociating reading processes on the basis of neuronal interactions. *J Cogn Neurosci* 17, 1753-1765.
- Milby, A.H., Halpern, C.H., Baltuch, G.H., 2008. Vagus nerve stimulation for epilepsy and depression. *Neurotherapeutics* 5, 75-85.
- Moran, R.J., Stephan, K.E., Kiebel, S.J., Rombach, N., O'Connor, W.T., Murphy, K.J., Reilly, R.B., Friston, K.J., 2008. Bayesian estimation of synaptic physiology from the spectral responses of neural masses. *Neuroimage* 42, 272-284.
- Nahas, Z., Teneback, C., Chae, J.H., Mu, Q., Molnar, C., Kozel, F.A., Walker, J., Anderson, B., Koola, J., Kose, S., Lomarev, M., Bohning, D.E., George, M.S., 2007. Serial vagus nerve stimulation functional MRI in treatment-resistant depression. *Neuropsychopharmacology* 32, 1649-1660.
- Narayanan, J.T., Watts, R., Haddad, N., Labar, D.R., Li, P.M., Filippi, C.G., 2002. Cerebral activation during vagus nerve stimulation: a functional MR study. *Epilepsia* 43, 1509-1514.
- Naritoku, D.K., Terry, W.J., Helfert, R.H., 1995. Regional induction of fos immunoreactivity in the brain by anticonvulsant stimulation of the vagus nerve. *Epilepsy Res* 22, 53-62.
- Osharina, V., Bagaev, V., Wallois, F., Larnicol, N., 2006. Autonomic response and Fos expression in the NTS following intermittent vagal stimulation: importance of pulse frequency. *Auton Neurosci* 126-127, 72-80.

- Paxinos, G., Watson, C., 1997. The rat brain in stereotaxic coordinates. Academic Press, Inc., San Diego.
- Penny, W.D., Stephan, K.E., Daunizeau, J., Rosa, M.J., Friston, K.J., Schofield, T.M., Leff, A.P., In press. Comparing families of dynamic causal models. *PLoS Comp Biol*.
- Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004. Comparing dynamic causal models. *Neuroimage* 22, 1157-1172.
- Roebroeck, A., Formisano, E., Goebel, R., 2009. The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. *Neuroimage*.
- Saper, C.B., 2002. The central autonomic nervous system: conscious visceral perception and autonomic pattern generation. *Annu Rev Neurosci* 25, 433-469.
- Saper, C.B., 2004. Central autonomic system. In: Paxinos, G. (Ed.), *The rat nervous system*. Elsevier Academic Press, San Diego, pp. 761-796.
- Schuyler, B., Ollinger, J.M., Oakes, T.R., Johnstone, T., Davidson, R.J., 2009. Dynamic Causal Modeling applied to fMRI data shows high reliability. *Neuroimage*.
- Schweinhart, P., Fransson, P., Olson, L., Spenger, C., Andersson, J.L., 2003. A template for spatial normalisation of MR images of the rat brain. *J.Neurosci.Methods* 129, 105-113.
- Siman-Tov, T., Mendelsohn, A., Schonberg, T., Avidan, G., Podlipsky, I., Pessoa, L., Gadoth, N., Ungerleider, L.G., Hendler, T., 2007. Bihemispheric leftward bias in a visuospatial attention-related network. *J Neurosci* 27, 11271-11278.
- Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J., Friston, K.J., 2009. Bayesian model selection for group studies. *Neuroimage* 46, 1004-1017.
- Sucholeiki, R., Alsaadi, T.M., Morris, G.L., 3rd, Ulmer, J.L., Biswal, B., Mueller, W.M., 2002. fMRI in patients implanted with a vagal nerve stimulator. *Seizure* 11, 157-162.
- Tracey, K.J., 2007. Physiology and immunology of the cholinergic antiinflammatory pathway. *J Clin Invest* 117, 289-296.
- Tracey, K.J., 2009. Reflex control of immunity. *Nat Rev Immunol* 9, 418-428.
- Van Der Zanden, E.P., Boeckxstaens, G.E., de Jonge, W.J., 2009. The vagus nerve as a modulator of intestinal inflammation. *Neurogastroenterol Motil* 21, 6-17.
- Weber, R., Ramos-Cabrer, P., Wiedermann, D., Van, C.N., Hoehn, M., 2006. A fully noninvasive and robust experimental protocol for longitudinal fMRI studies in the rat. *Neuroimage* 29, 1303-1310.
- Wise, R.G., Ide, K., Poulin, M.J., Tracey, I., 2004. Resting fluctuations in arterial carbon dioxide induce significant low frequency variations in BOLD signal. *Neuroimage* 21, 1652-1664.
- Xu, F., Frazier, D.T., 1995. Medullary respiratory neuronal activity modulated by stimulation of the fastigial nucleus of the cerebellum. *Brain Res* 705, 53-64.

Table Legends

Table 1: List of ROIs used in DCM.

See Figure 1. Coordinates indicate the centre of ROIs, in the atlas of Paxinos and Watson referenced to bregma (Paxinos and Watson, 1997). D-V: dorso-ventral; M-L: medio-lateral; A-P: antero-posterior.

DCM ROI	Anatomical structures	Abbreviations in Paxinos and Watson atlas	Centre coordinates		
			D-V	M-L	A-P
AI	Nucleus of the solitatory tract / Vestibular nucleus	Sol / VE	-8	±1.2	-13.2
CA1	Field CA1 of hippocampus	CA1	-3	±1.2	-2.8
Cg	Cingulate cortex	Cg1 / Cg 2	-2.4	±0.3	0.5
NTS	Nucleus of the solitatory tract / Vestibular nucleus / Prepositus nucleus / Intercalated nucleus of the medulla / Gracile nucleus	Sol / VE / Pr / In / Gr	-8	±1.2	-13.2
RSG	Granular and agranular retrosplenial cortex	RSGa / RSGb / RSA	-1.7	±0.8	-5.6

Figures

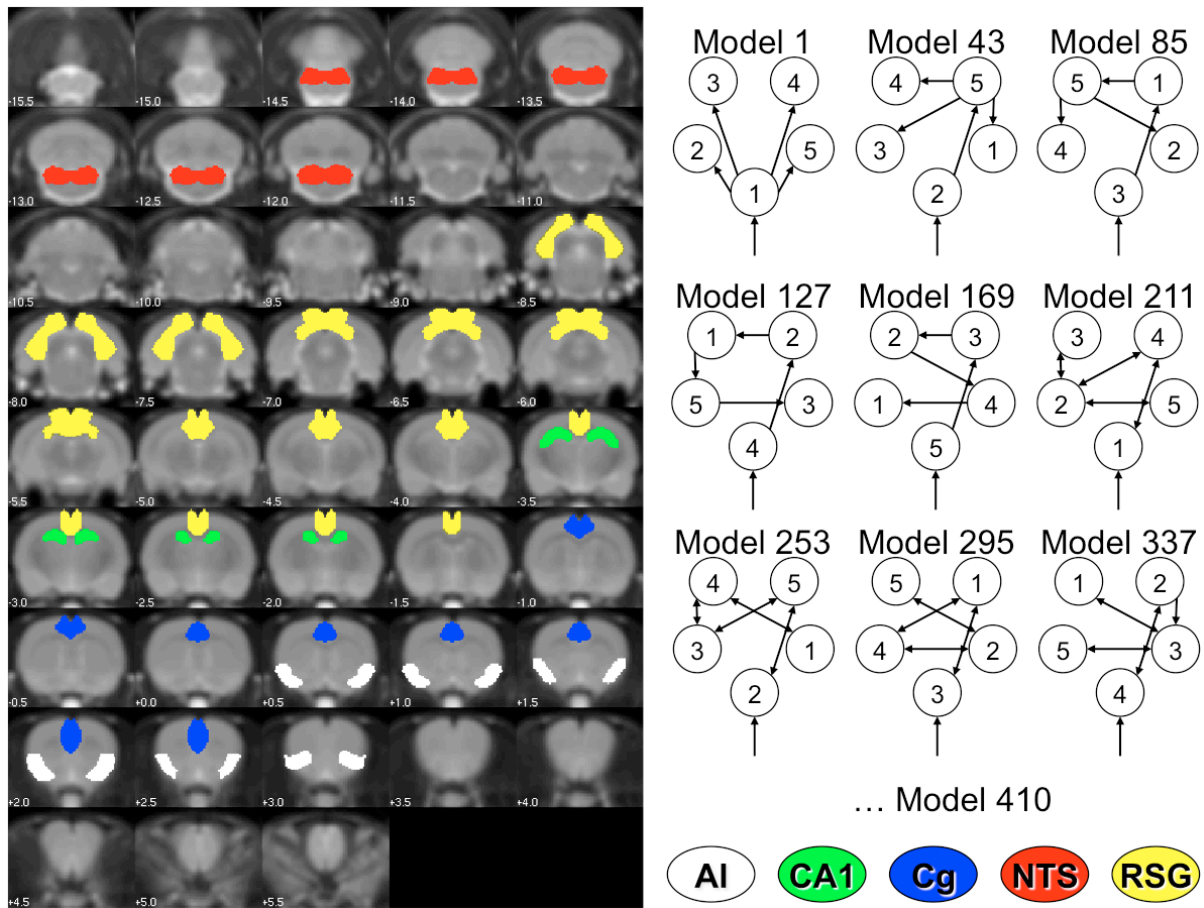


Figure 1: Dynamic Causal Modelling of VNS in rat. Left: Regions of interest are superimposed on a T2 rat template (coronal slices shown every 0.5 mm, in posterior to anterior order). Right: Nine among the 410 tested DCMs of VNS. NTS: nucleus tractus solitarius; CA1: hippocampus (anterior part); AI: insular cortex; RSG: retrosplenial granular cortex; Cg: cingulate cortex.

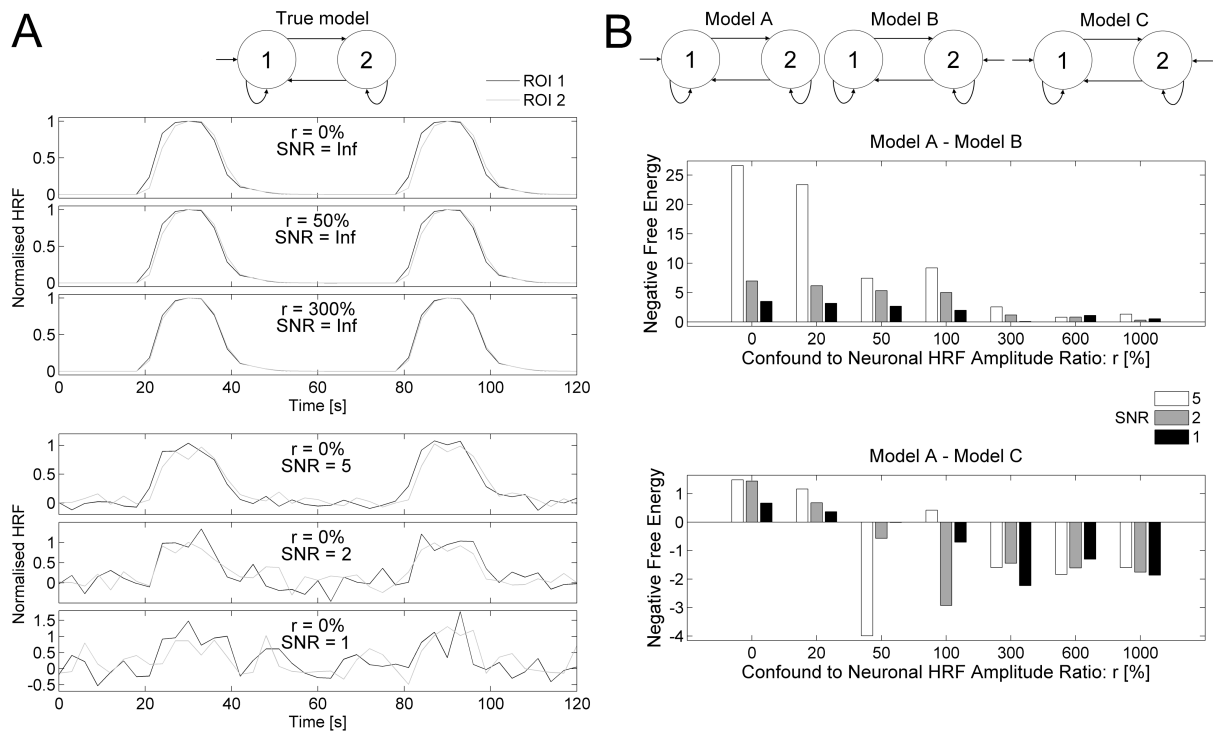


Figure 2: Simulation of Dynamic Causal Modelling with correlated confounds. A: 2 minutes of generated time series (HRF with amplitude normalised to 1) with the “true” model defined by Equation (1). The three first panels show different degrees of contamination by confounds (common mode of original time series) defined by different values of parameter r , without random noise ($SNR = Inf$, *i.e.* $\sigma = 0$). The three last panels show examples of simulated BOLD time series for the different values of SNR. B: Difference of negative free energy (approximate of log model evidence) between Models A & B (top) and Models A & C (bottom), when fitted to data shown in A and averaged over 50 realisations, as a function of parameter r and SNR .

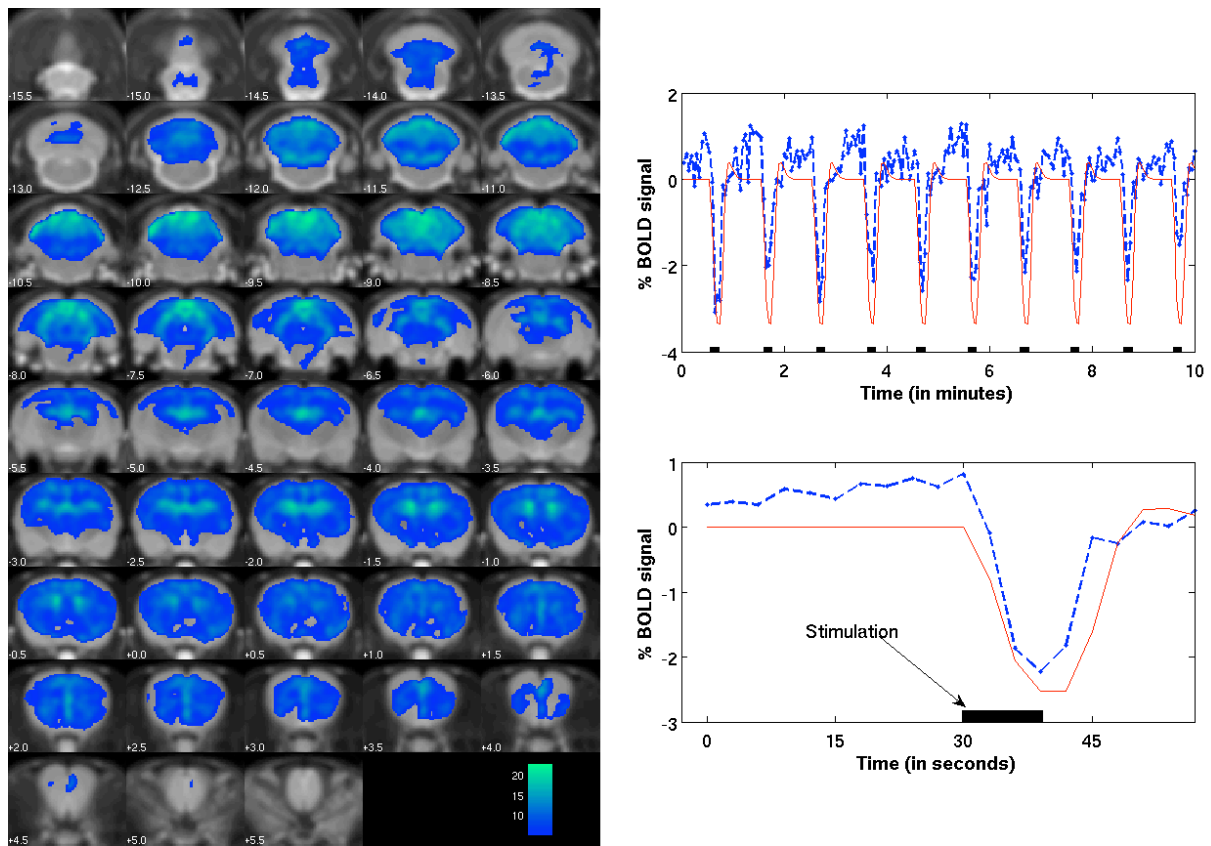


Figure 3: Deactivations induced by vagal nerve stimulation (group study, $n=8$). Left: Map of negative t-value (thresholded at $p<0.001$, FWE corrected) showing deactivations. Top right: group average of BOLD signal variations in the NTS (blue) superimposed on the regressor of interest (red) obtained after convolution of the canonical hemodynamic response with the stimulus function (black). Bottom right: Average over all stimulations.

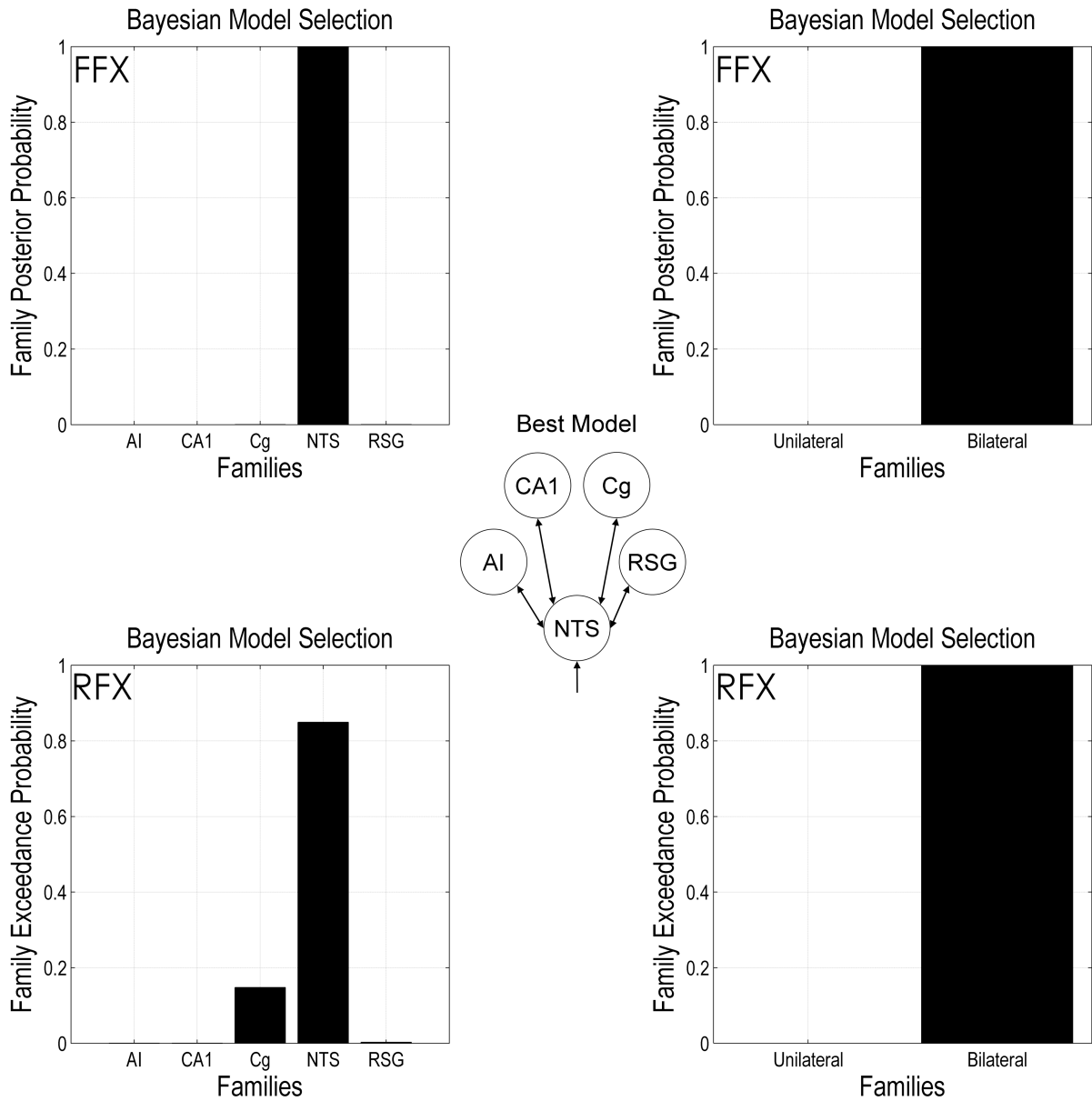


Figure 4: Group level (n=5) Bayesian selection of families of the 410 tested models. Top: fixed-effect analysis (FFX) showing family posterior probability; Bottom: random-effect analysis (RFX) showing family exceedance probability. Left: Families were constructed to estimate the input station of the VNS (family labels indicate the input structure) and results indicate the NTS as the most likely. Cg comes second and probability is very weak for other scenarios. Right: Families grouped together models with unidirectional or bidirectional interregional connectivity. Clear evidence is for the existence of bidirectional long range neuronal connections. Middle: Most likely model amongst the 410 tested models.