Neuropsychologia 99 (2017) 1–11

Contents lists available at ScienceDirect

Neuropsychologia

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

Directed network discovery with dynamic network modelling

Stefano Anzellotti[a,*], Dorit Kliemann[a], Nir Jacoby[b], Rebecca Saxe[a]

[a] MIT, United States
[b] Columbia University, United States

ARTICLE INFO

Keywords:
Connectivity
Emotions
Dynamic network modelling
Granger causality
Dynamic causal modelling

ABSTRACT

Cognitive tasks recruit multiple brain regions. Understanding how these regions influence each other (the network structure) is an important step to characterize the neural basis of cognitive processes. Often, limited evidence is available to restrict the range of hypotheses a priori, and techniques that sift efficiently through a large number of possible network structures are needed (network discovery). This article introduces a novel modelling technique for network discovery (Dynamic Network Modelling or DNM) that builds on ideas from Granger Causality and Dynamic Causal Modelling introducing three key changes: (1) efficient network discovery is implemented with statistical tests on the consistency of model parameters across participants, (2) the tests take into account the magnitude and sign of each influence, and (3) variance explained in independent data is used as an absolute (rather than relative) measure of the quality of the network model. In this article, we outline the functioning of DNM, we validate DNM in simulated data for which the ground truth is known, and we report an example of its application to the investigation of influences between regions during emotion recognition, revealing top-down influences from brain regions encoding abstract representations of emotions (medial prefrontal cortex and superior temporal sulcus) onto regions engaged in the perceptual analysis of facial expressions (occipital face area and fusiform face area) when participants are asked to switch between reporting the emotional valence and the age of a face.

New and noteworthy

In this article we introduce a new analysis method (Dynamic Network Modelling or DNM) which performs efficient for network discovery by testing the consistency of vector autoregressive (VAR) model parameters across participants. DNM provides information about the direction and sign (inhibitory vs excitatory) of influences between brain regions, and generates measures of variance explained in independent data to evaluate quality of fit. The method is applied to brain regions engaged in emotion recognition, individuating a similar network structure across two separate experiments.

1. Introduction

When we perform a task, such as recognizing a face, attributing mental states to others, or understanding a sentence, multiple brain regions are engaged (Ishai, 2008; Gallagher and Frith, 2003; Fedorenko and Thompson-Schill, 2014). Studying how these brain regions influence each other is an important step to understand the neural mechanisms underlying task performance. Influences between brain regions can be specific to the particular task a participant is performing. For example, face-selective brain regions might influence each other more when we recognize a face than when we recognize a scene. For this reason, we need a method that goes beyond measuring the presence of anatomical connections between regions, and to investigate the relations between the regions' responses in the context of a specific experimental paradigm.

The direction of an influence can convey information about its function. For example, an influence from the ventral visual stream to prefrontal cortex is likely to convey bottom-up perceptual information to categorization and decision processes, while an influence from prefrontal cortex to the ventral visual stream is more likely to affect visual processing via top-down attentional selection (Buschman and Miller, 2007). Directed influences between brain regions can also contribute to characterize the functional role of a brain region by investigating how it receives inputs and conveys outputs to other regions with functional roles that are better understood.

Directed influences can be studied using temporal precedence: observing if earlier responses in a region contribute to predicting later responses in another region. Studying temporal precedence with functional magnetic resonance imaging (fMRI) presents unique advantages but also unique challenges. Data can be acquired noninva-

* Correspondence to: MIT, 77 Massachusetts Avenue, 46-4017, Cambridge, 02139 MA, United States.
E-mail address: anzellot@mit.edu (S. Anzellotti).

http://dx.doi.org/10.1016/j.neuropsychologia.2017.02.006
Received 9 September 2016; Received in revised form 18 January 2017; Accepted 3 February 2017
Available online 16 February 2017
0028-3932/

sively, with good resolution, and covering the entire brain, making fMRI well-suited to study long-range influences and investigate uniquely human aspects of cognition. At the same time, fMRI measures Blood-Oxygen Level Dependent (BOLD) signal, whose timing is affected by the local properties of vasculature. Adequate steps must be taken to control for the variability in BOLD timing between regions.

Depending on the evidence already available, different approaches to studying influences can be more or less suitable. In some cases, the previous evidence can be used to restrict a-priori the hypotheses about the influences between a set of brain regions, paving the way for confirmatory analyses. Often, however, limited evidence is available, and a very broad range of different influences are possible. Currently, the main techniques used to study influences between brain regions with fMRI are Granger Causality (GC, Roebroeck et al., 2005) and Dynamic Causal Modelling (DCM, Friston et al., 2003). Each of these techniques has important strengths, but also properties that may not be desirable in some data analysis contexts.

1.1. Granger causality

In GC (Roebroeck et al., 2005), the influence of one brain region on another is measured as a function of the variance in the responses in the latter region that is explained by earlier responses in the former, in addition to the variance explained by the latter region itself (see Appendix B for a more formal description). GC thus offers an intuitive measure of influences between regions which is not computationally costly to obtain. Nevertheless, GC has some disadvantages when applied to haemodynamic responses measured by fMRI. First, in its current form Granger causality is difficult to apply to the modelling of influences in different conditions within fast event-related designs. Separate models are used for the different conditions that need to be compared, and in fast event-related designs this would require breaking up the timeseries in short chunks that would impair autoregressive modelling. Second, since standard GC is based on variance explained, it does not measure whether stronger responses in one region lead to stronger or weaker responses in another (i.e. the 'sign' of the influence), or even whether this feature of the influence is consistent across participants. Investigating the sign of interactions provides additional insights into their functional role, and its interpretation in terms of excitation and inhibition has been pioneered in the clinical literature in recent studies (Hamilton et al., 2011; Chen et al., 2009).

1.2. Dynamic causal modelling

In DCM (Friston et al., 2003), the change in neural response in each brain region is modeled as a function of the stimulus and the input from other regions using ordinary differential equations (see Appendix C for a more formal description). Given a set of brain regions and a set of conditions, there is a fixed number of possible parameters. A candidate model is specified by providing, for each possible parameter, a 1 if that parameter will be included in the candidate model (whether a connection'exists' in that candidate model), and a 0 otherwise. The parameters that are included are then estimated with the expectation-maximization (EM) algorithm, and the best candidate model is chosen using Bayesian model comparison. The presence of condition-dependent influences makes DCM suitable for the analysis of fast event-related designs, and the use of priors mitigates overfitting acting as a form of regularization.

The proposed DNM approach adapts many of the strengths of DCM, but takes a different approach to two key challenges. The first challenge is network discovery. DCM is designed as a confirmatory technique, and therefore it is particularly suitable for choosing between models that have been identified on the basis of prior evidence. Searching through larger hypothesis spaces in DCM is computationally costly (but see Friston and Penny (2011) for an ingenious technique to increase speed); but more importantly, DCM computes the posterior probability of the best model, relative to the set of considered models. This estimate is most useful when researchers can be confident that their a priori hypothesis space includes most or all plausible models of the network. By contrast, DNM is designed as an exploratory or network-discovery approach. We propose that DNM can be used by researchers who cannot restrict their hypothesis space to a small set of models based on existing evidence, or who are unsure whether the limited temporal resolution of fMRI data is sufficient to reveal any robust inter-regional influences.

Whereas DCM compares full network models (each model is a description of the overall structure of the network), DNM assesses the variance in independent data explained by each individual parameter (or connection), using random effects statistical tests on parameter values to efficiently search all possible connections. A consequence of this difference is that whereas the best model chosen by DCM includes both the existing and the absent connections, DNM (like traditional null-hypothesis tests) makes claims about the existing connections, but not about the absent connections. On the other hand, DNM provides an absolute, not relative, estimate of the variance explained by each connection in independent data. This intuitive measure of the quality of the model is especially useful for exploratory or network-discovery analyses.

Existing packages in DCM employ free energy to assess quality of the model fitting to the data. Although this function is based in the same data used for model selection, it is not biased in favor of complex models. DCM also offers the option to calculate variance explained by each model, but using the same data used for model selection. As a consequence, the variance explained by the selected model is over-estimated. By contrast, DNM uses variance explained in independent data. Similarly to free energy, this is not biased by model complexity. However, variance explained in independent data has an important additional asset: it is an accurate estimate of the absolute measure of goodness of fit. Variance explained in independent data can be used to how well the model fits the current data (for example, making clear when even the best model provides a relatively poor fit), but also to compare model fits across experiments and populations. Therefore we chose variance explained in independent data as a key measure to evaluate quality of fit in DNM.

The second challenge is how to combine evidence across participants who may have individual differences in the strength of each connection. In DCM, each candidate model specifies the existence and direction of connections, but not their value (Stephan et al., 2010). That is, in DCM, a connection is deemed present if it explains variance, regardless of whether the parameters are similar (or even the same sign, that is 'excitatory' vs 'inhibitory') across participants. This approach grants DCM the flexibility to identify connections with highly variable strengths across participants, buthas consequences for the intuitive interpretation of the resulting model graphs. A high degree of variability in the parameter values across participants does not affect the probability of the model given the data. By contrast, DNM follows traditional null-hypothesis testing, in assessing the reliability of the magnitude (and sign) of each parameter across participants. This difference can be illustrated using methods for testing whether an experimental condition leads to a significantly greater response than baseline. In standard fMRI analyses, using a general linear model, a voxel is deemed to show a significant response if the magnitude (i.e. beta parameter) of response is reliable (similar in both magnitude and sign) across participants. A valid but different analysis would ask whether including the predictor for the experimental condition improves the fit of the model across participants (enough to compensate for the increased complexity of the model). In this case, the beta values might be highly variable, or even positive for some participants and negative for other participants, as long as they contribute enough to improving the model fit in each participant. Both of these analyses are valid, but address different questions. DCM takes an approach similar to the second analysis: when a model including is selected, we can infer

that each included connection improved the quality of fit enough to compensate for the additional complexity, but the parameter value for that connection might be quite different across participants. This observation has been recently raised in the MEG literature using DCM (Pinotsis et al., 2016; Friston et al., 2016), where it has been addressed using an approach based on hierarchical Bayesian models. In DNM, we adopt an alternative solution: using random effect statistical tests on the parameter values. This solution is intuitive and lends itself to regularization-based solutions to perform network discovery within large networks.

1.3. Dynamic network modelling

Given the considerations outlined in the previous sections, we set out to develop a new conservative method for exploratory analysis to model influences between brain regions in fMRI, that would meet a set of criteria: (1) control for the differences in the shape of haemodynamic responses in different regions, (2) provide a computationally efficient method for network discovery, (3) provide rigorous estimates of the magnitude and the sign of influences (whether they are inhibitory or excitatory), and (4) offer absolute measures of goodness of fit (rather than relative to other models) in independent data. To satisfy these desiderata, here we propose a novel approach to modelling directed influences between brain regions that builds on insights from GC and DCM and introduces some new ideas. We begin with deconvolution as in DCM, and proceed with a regularized vector-autoregressive (VAR) modelling, obtaining a procedure that we refer to as Dynamic Network Modelling (DNM). This two-step procedure in which DCM-like deconvolution is followed by subsequent network modelling has been shown to successfully control for HRF shape (David et al., 2008). Like GC, DNM models the effect on a region of the stimuli and of other regions with a system of linear equations. Like DCM, DNM includes parameters for fixed and condition-dependent connections as well as driving effects of the experimental conditions, and it can model independent connections in opposite directions between the same brain regions. Unlike both, DNM performs efficient network discovery with random effects tests on the stability of VAR model parameters across participants. DNM is validated with computer simulations, comparing an approach in which all parameters are fit simultaneously, with an approach in which autoregressive parameters, condition-dependent effects and inter-regional influences are fit sequentially in a hierarchical model. The hierarchical model can be used to generate conservative estimates of the additional variance explained in independent data by the inter-regional influences.

1.4. Empirical application

We adopted DNM to investigate the influences between brain regions involved in the recognition of emotional valence: an ideal case study because recent studies (Peelen et al., 2010; Skerry and Saxe, 2014) identified multiple brain regions encoding information about valence, and because recognition of emotional valence is likely to engage species-specific processes whose investigation necessitates noninvasive measurement. A brief glance at a person's face provides rich information: about the person's identity, age and gender, and also about their current emotional experience. Many brain regions are involved in extracting this information (Anzellotti and Caramazza, 2015). For example, the occipital and fusiform face areas (OFA and FFA) respond selectively to faces (Sergent et al., 1992; Kanwisher et al., 1997; Gauthier et al., 2000) and encode information about many facial features, including the valence of the emotions communicated by facial expressions (Peelen et al., 2010; Furl et al., 2012; Skerry and Saxe, 2014). The posterior superior temporal sulcus (pSTS) and medial prefrontal cortex (MPFC) are sensitive to cues about another person's emotional experience, whether conveyed in facial expressions or in body movements, vocal tones, or even abstract descriptions of events (Narumoto et al., 2001; Winston et al., 2005; Fusar-Poli et al., 2009; Peelen et al., 2010; Furl et al., 2012; Skerry and Saxe, 2014).

A key open question concerns the interaction between these two kinds of information (see Ishai, 2008; Bressler and Menon, 2010): how does the facial form processing in OFA and FFA interact with the more invariant representations of emotional valence in pSTS and MPFC? One possibility is that recognition of emotion in facial expressions is accomplished mainly by successive bottom-up feature extraction, building increasingly invariant representations at each stage. If so, one would expect influences mainly from OFA and FFA to pSTS and MPFC. Another possibility, however, is that higher-level representations of emotions influence the processing of the facial form. If so, one might expect that inter-regional influences include, or are even dominated by, top-down influences from pSTS and MPFC on the processing in OFA and FFA.

Structural connectivity analyses suggest that although OFA and FFA are connected, there are no direct (i.e. monosynaptic) tracts connecting pSTS to the OFA or FFA (Davies-Thompson and Andrews, 2011; Ethofer et al., 2011, 2013; Gschwind et al., 2011; Pyles et al., 2013). Direct anatomical connections between MPFC and OFA/FFA are also unlikely. Nevertheless, the magnitude of spontaneous activity during rest (i.e. in the absence of any emotional stimuli) is correlated between FFA and pSTS (Turk-Browne et al., 2010) suggesting at least indirect connectivity between these regions.

We applied DNM to data from two independent experiments in which participants viewed brief (4 s) movies depicting a stranger's positive or negative emotional experiences. Multi-voxel pattern analyses of the first dataset suggested that information about the valence of these expressions is present in OFA, FFA, pSTS and MPFC (Skerry and Saxe, 2014). We therefore used these data, without a-priori restrictions on the connectivity profiles, to ask (1) how these regions interact during the recognition of emotions, and (2) whether DNM explains additional variance in independent data (as compared to a standard GLM) and produces reliable results (replicating connectivity profiles across different datasets).

2. Materials and methods

2.1. Preprocessing and deconvolution

Given a set of n ROIs, the mean BOLD signal is extracted from each of the ROIs, and detrended with the SPM function 'spm_detrend' (http://www.fil.ion.ucl.ac.uk/spm/). Model-based deconvolution was performed as in David et al. (2008), using DCM software based on a modification of the balloon model to obtain estimates of the shape of the haemodynamic response function (HRF; Friston et al., 2003). This deconvolution method was shown to control successfully for variability in the HRF shape between different brain regions in fMRI data, leading to fMRI-derived estimates of connectivity that correspond to those obtained with intracortical EEG (David et al., 2008). For this reason, we do not provide additional validation tests here for this deconvolution approach.

2.2. Interaction model

Deconvolution yields a deconvolved timeseries $z_i(t)$ for each ROI i=1,...,n. The vector of deconvolved timeseries $\mathbf{z}(t) = [z_1(t); \ldots; z_n(t)]$ is modeled with a bilinear vector autoregressive model. A bilinear model can model condition-dependent effects even in the context of fast event-related designs, and it does not require integration to estimate parameters thus reducing the computational costs. We have therefore

$$\mathbf{z}(t) = \sum_{k=1}^{l} A(k)\mathbf{z}(t-k) + \sum_{k=1}^{l}\sum_{j=1}^{m} u_j(t-k)B_j(k)\mathbf{z}(t-k) + C\mathbf{u}(t) \quad (1)$$

where k=1,...,l are the time lags between predicted and predictor

responses, the matrices $A(k)$ contain parameters for fixed influences at lag k, $u_j(t)$ is 1 if condition j is presented at time t and 0 otherwise, the matrices $B_j(k)$ contain parameters for condition-dependent influences in condition j at lag k, and C contains parameters for the effect of the conditions $\mathbf{u}(t)$.

2.3. Model fitting

Two fundamental considerations were made in the context of model fitting. First, fitting all model parameters in a single stage could lead any shared variance between parameters for the influence of the regions on themselves (autoregressive parameters) and parameters for the influences between different regions to be split equally among them. Timecourses of fMRI data show a high degree of temporal autocorrelation (Woolrich et al., 2001), and fitting all parameters in a single stage could attribute part of the variance explained by autocorrelation to influences between regions, and this might result in exceedingly liberal estimates of the influences between regions. For this reason, DNM in which model parameters were fit simultaneously was compared to DNM with a hierarchical fitting procedure, in which in a first stage the autoregressive parameters are estimated, subsequently the effect of conditions, and in the end the parameters for the influences between different regions. The hierarchical fitting procedure is made explicit in the following set of equations:

$$\begin{aligned}\mathbf{z}(t) &= \sum_{k=1}^{l} \mathrm{diag}(A(k))\mathbf{z}(t-k) + \boldsymbol{\epsilon}^{I}(t)\boldsymbol{\epsilon}^{I}(t) = C\mathbf{u}(t) + \boldsymbol{\epsilon}^{II}(t)\boldsymbol{\epsilon}^{II}(t) \\ &= \sum_{k=1}^{l} (A(k) - \mathrm{diag}(A(k)))\boldsymbol{\epsilon}^{II}(t-k) + \sum_{k=1}^{l}\sum_{j=1}^{m} u_j(t-k)B_j(k)\boldsymbol{\epsilon}^{II}(t \\ &\quad - k) + \boldsymbol{\epsilon}^{III}(t) \qquad (2)\end{aligned}$$

where $\epsilon^{I}(t) = [\epsilon_1^{I}(t), \ldots, \epsilon_n^{I}(t)]$, $\epsilon^{II}(t) = [\epsilon_1^{II}(t), \ldots, \epsilon_n^{II}(t)]$, and $\epsilon^{III}(t) = [\epsilon_1^{III}(t), \ldots, \epsilon_n^{III}(t)]$ are the residuals for the three modelling stages and the n ROIs. The first equation implements an autoregressive model, the second equation a linear model with conditions as predictors, and the final equation models the remaining variance as a function of the influences between different brain regions.

The models' time lag k can be determined using the Akaike Information Criterion (AIC), but in practice given the high number of parameters the selected k is equal to 1. One potential cost of DNM's fitting procedure is that estimates of the shape of the HRF are generated before fitting the influence parameters rather than simultaneously. However, previous studies found that the HRF could be estimated accurately even with such a sequential procedure (David et al., 2008). Models were fit with ordinary least squares (OLS), regularization can be used when considering large networks.

3. Validation

The performance of the method was validated with computer simulations, by defining VAR processes and using them to generate data with added noise. We tested DNM's ability to accurately infer the model parameters and to individuate influences that are stable across participants. The deconvolution approach has been validated extensively in previous work (Friston et al., 2003; David et al., 2008), so the present validation focused on the VAR modelling component.

Data were generated using models of the following form:

$$\mathbf{z}(t) = A\mathbf{z}(t-1) + C\mathbf{u}(t).$$

For each simulation iteration, we first generated group-level matrices. The group-level matrix A_g was generated extracting diagonal parameters from a normal distribution of mean 0.5 and standard deviation 0.1 (reflecting the high autocorrelation between a region and itself) and off-diagonal parameters from a normal distribution of mean 0 and standard deviation 0.1. The group level matrix C_g was generated extracting each parameter from a normal distribution with mean 0.5 and standard deviation 0.2. At a second stage, individual subject matrices $(A_s1, \ldots, A_sn)$ were generated from normal distributions having as mean for each parameter the group-level value for that parameter. Stability of the simulated VAR processes was ensured by checking that the eigenvalues of the A_{si} matrices were all smaller than 1. For each choice of number of subjects n from 10 to 20, of number of regions m from 3 to 5, we generated 10 different group-level models and performed DNM analysis with both the simultaneous and the hierarchical approaches. The initial timepoint $\mathbf{z}(1)$ was extracted from a multivariate uniform distribution on the hypercube $[0, 1]^m$, and zero-mean gaussian noise was added to the data. Multiple levels of noise were tested, with standard deviations ranging from 0.1 to 1 respectively. For each simulation, we assessed the accuracy with which DNM could model the influences between regions, evaluating significance for each influence with a t-test over DNM model parameters across participants, and counting a significant inference for influence from region k to region j as accurate if (1) it corresponded to a true influence between regions that is stable across participants (the t-test performed on the true subject values $A_{si}(j, k)$ was significant) and (2) the sign of the interaction was accurately inferred. In addition to the accuracy, we calculated the proportion of true influences that went undetected by the model.

3.1. Simultaneous and hierarchical model fitting

As expected, the number of undetected true influences was higher in the hierarchical model than in the simultaneous model (Fig. 1A). However, importantly, this did not result in a lower accuracy of the simultaneous model, and in fact, the simultaneous model was overall more accurate than the hierarchical model across different numbers of participants and regions, and across different levels of noise (Fig. 1B). For this reason, in the analysis of real fMRI data we will use the simultaneous model fitting procedure to estimate influences between regions, and we will restrict the hierarchical fitting procedure to the purpose of estimating conservatively the additional variance explained by the inter-regional influences. Accuracy for the simultaneous model was very high across all noise levels, validating DNM as a technique to infer the presence and sign of influences between brain regions.

3.2. Comparison with Granger causality

As an additional test of DNM's ability of taking into account variability across participants in the sign of influences, we compared DNM with GC (as implemented) on two sets of 1000 simulated datasets with 20 participants and 3 regions. The first 1000 datasets were generated as described in the validation analysis. In the remaining 1000, instead, the parameter values for off-diagonal entries of $A_{si}(j, k)$ in individual subjects were extracted from normal distributions having as mean $A(j, k)$ for even-numbered subjects, but $-A(j, k)$ for odd-numbered subjects. Since Granger Causality does not take into account the sign of influences, we expected it to perform well when participants had consistent signs, but to report consistent influences even in the presence of high intersubject variability in the sign of the parameters. This does not imply that the GC results are incorrect, but that great caution needs to be exerted when interpreting them: consistent GC across participants indicates that an influence plays a 'similarly important' role across participants, but the actual role of the influence might vary widely across participants, being excitatory in some and inhibitory in others. We will refer to inferences of significant influences when the t statistic calculated with the true values $A_{si}(j, k)$ was non significant as 'false positives' – but it is important to emphasize that in the context of GC this set of inferences includes in part cases that may be validly interpreted to indicate significant unsigned 'importance' of an influence. Since GC is always positive, we could not use t-tests to assess significance, therefore we estimated the distribution of Granger

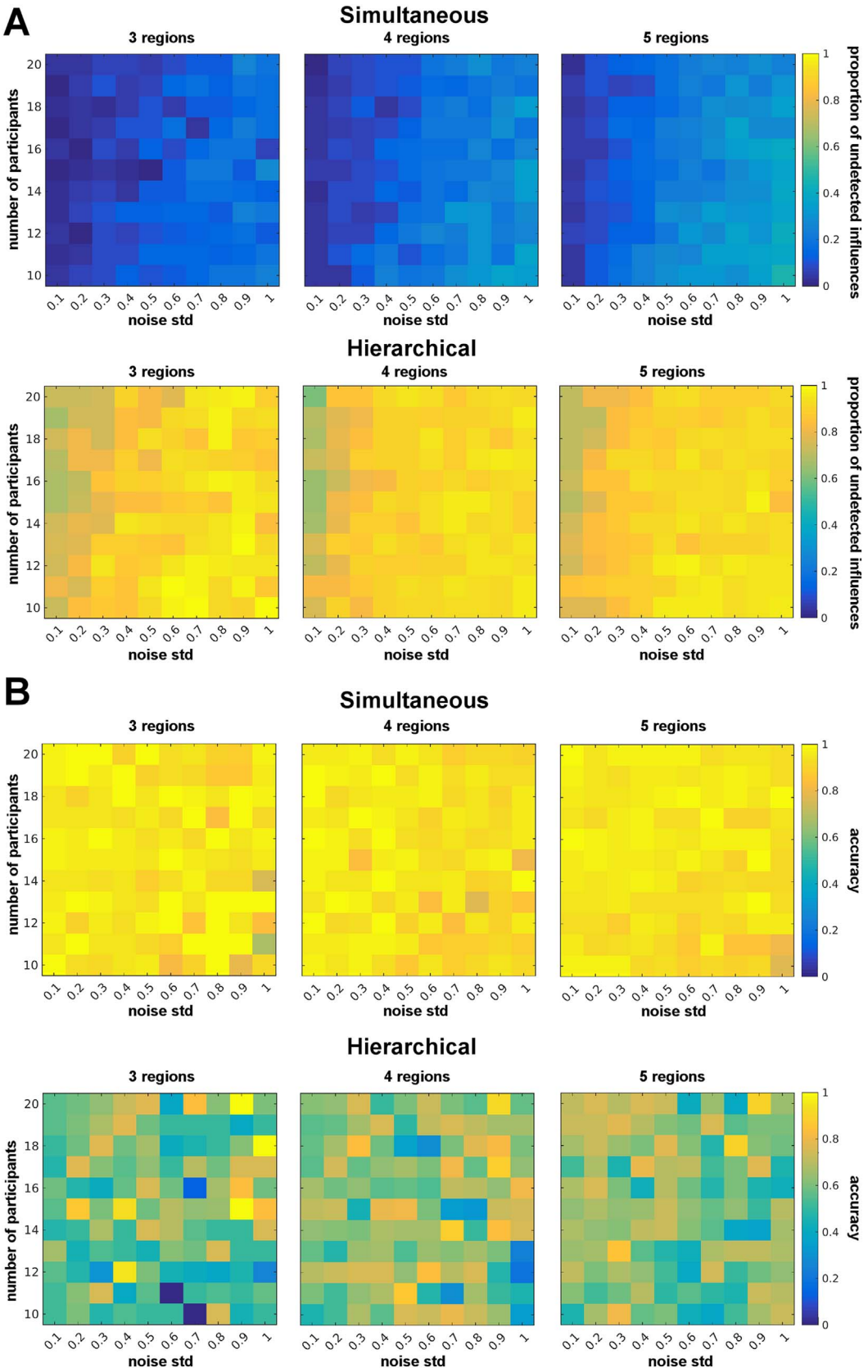

Fig. 1. (A) proportion of undetected influences and (B) accuracy at inferring the presence and sign of the influences for the simultaneous models and hierarchical models.

Causality F values under the null hypothesis using an additional 1000 simulations in which the A_{si} matrices were generated independently for each subject. In the subsequent analysis, an influence was considered significant if it was strictly within the top 5% of the maxima of the connectivity matrices obtained under the null hypothesis. We set the thresholds for DNM using the same simulation-based approach and

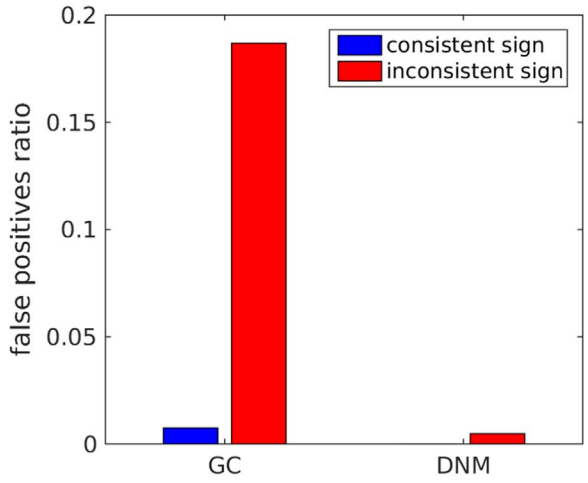

Fig. 2. Ratio of 'false positives' for GC and DNM when participants have influences of the same sign (blue bars) and when participants have high variability in the sign of influences (red bars). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

using the same set of simulations, to ensure that different numbers of false positives would not result from different approaches to determine significance. When participants had consistent signs, both GC and DNM showed a very small number of false positives (less than 1% of the total number of connections: 45 for GC and 0 for DNM, Fig. 2 A). However, when participants had different signs, GC showed a large increase in the number of false positives (1121, that is 18.68% of the total number of connections), while DNM did not (14, that is 0.48% of the total number of connections Fig. 2B, note the scale of the y axis).

4. Experimental materials and procedures

Two different experiments testing emotion recognition from visual stimuli were completed with two different groups of participants. In each experiment, videos prompting the attribution of an emotion to a target were shown to participants while fMRI data were collected. The blood-oxygen level dependent (BOLD) timeseries measured were used to model connectivity between face selective regions and regions encoding information about emotional valence (OFA, FFA, pSTS and MPFC, see Skerry and Saxe, 2014; Kliemann et al., 2016).

4.1. Experiment 1: stimuli and task

A total of 26 volunteers took part in experiment 1 (all right handed, n=10 female, age range: 19–44, mean=26.25, SD=6.12). All participants had normal or corrected to normal vision and no history of neurological or psychiatric disorders. Participants gave informed consent in written form in line with the requirements of MIT's institutional review board.

The experiment consisted of 8 runs, a face localizer, and a theory of mind localizer (Dodell-Feder et al., 2011; stimuli are available at http://saxelab.mit.edu/superloc.php). The theory of mind localizer was not analyzed for this article. In the 8 runs, participant viewed videos of faces expressing an emotion ('expressions condition') and of simple geometric characters experiencing an event eliciting emotions ('situations condition'). Participants were asked to rate on a 1–4 scale after each video the intensity of the emotion experienced by the characters. Face stimuli involved a close-perspective view on single entity, therefore they were presented at 7.8×7.4° of visual angle, while the context animations were presented at 16.7×12.5°. Videos of facial expressions were obtained from movies, producing a set of relatively naturalistic stimuli, achieving a balance between external validity (see Zaki and Ochsner, 2009; Spunt and Lieberman, 2012) and experimental control. Each trial consisted of the presentation of a 4 s video followed by a 1.75 s response screen and a 250 ms blank. A fixation cross of variable duration (0–14 s) was presented between trials.

In the face localizer, participants viewed videos of children's faces and of moving objects (from Pitcher et al., 2011). 30 videos were shown for each condition, grouped in blocks of 6. Participants had to perform a 1-back task detecting repetitions of identical videos. Each video lasted 3 s, and was followed by a blank screen presented for 333 ms, for a total block duration of 20 s. A 2 s blank was shown between every two blocks, and 12 s blocks of fixation were included at the beginning, middle and end of the run. Participants were shown 2 localizer runs, each lasting 5 min. The order of conditions was counterbalanced within runs, across runs, and across participants.

4.2. Experiment 2: stimuli and task

A total of 28 volunteers took part to experiment 2 (all right handed, 11 female, age range: 21–33, mean=26.6, SD=4.2). All participants had normal or corrected to normal vision and no history of neurological or psychiatric disorders. Participants gave informed consent in written form in line with the requirements of MIT's institutional review board. The experiment was composed of 8 functional runs (lasting 372 s each) and a face localizer (184 s) (Hariri et al., 2000).

In the experimental runs, the facial expression videos from Experiment 1 were presented. Unlike in Experiment 1, however, participants were asked to perform two different judgements on the facial expression videos. In half of the trials, they were asked to assess the valence of the emotion (positive or negative), in the remaining trials they were asked to assess the age of the person in the video (younger than 40, older than 40). Each trial consisted of the presentation of a screen indicating the task to be performed (1 s), a blank (4–12 s, mean=8 s), a video (4 s), an additional blank (250 ms) and a screen showing a plus and a minus arranged horizontally as a response cue (1.75 s). Participants pressed the left or right button to select the plus (for positive emotions or for faces of people over 40) or the minus (for negative emotions or for faces of people under 40). At the end of each run, a 12 s blank screen was presented. Participants were excluded if they performed below 83% accuracy in two or more runs, a criterion established before starting fMRI analysis. Data from three participants (1 female) were excluded due to this exclusion criterion.

The face localizer consisted in the presentation of 30 s blocks of geometric shapes and blocks of faces. Each block consisted of 15 2 s trials in which participant had to match the expression of a face (or the shape) in the upper part of the screen with one of two expressions (or shapes) in the lower part of the screen.

4.3. Data acquisition

Data acquisition parameters were identical in the two experiments. Data were acquired with a 3T Siemens Tim Trio scanner in the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT, using a Siemens 32-channel phased array head coil. A high-resolution (1 mm isotropic voxels) T-1 weighted MPRAGE anatomical scan was followed by functional scans using a gradient-echo EPI sequence sensitive to blood-oxygen-dependent (BOLD) contrast (repetition time [TR]=2 s, echo time [TE]=30 ms, flip angle=90°, voxel size 3×3×3 mm). 32 axial slices of 64×64 voxels aligned with the anterior/posterior commissure were acquired in each volume, covering the whole-brain except the cerebellum.

4.4. ROI definition and interaction modelling

Data from the face localizers were preprocessed using SPM8 running on MATLAB R2010b and custom in-house software, and were

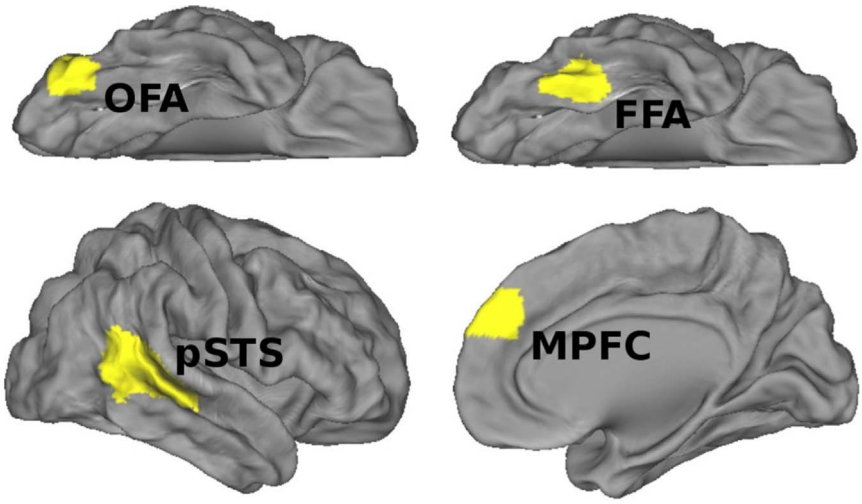

Fig. 3. Group-level search spaces for the regions of interest.

modeled in both experiments with a standard GLM with regressors for each stimulus category. Regions of interest for the occipital face area (OFA), fusiform face area (FFA) and posterior superior temporal sulcus (pSTS) were defined in individual participants using group-defined search spaces (Fig. 3). 1 of the 26 participants in Experiment 1 did not complete the face localizer, in one OFA and FFA could not be localized, and in 1 pSTS could not be localized. In these cases ROIs were derived from group-level activation maps obtained from the localizers of the other participants. In addition to these face-selective ROIs, a region of interest was defined for MPFC at the group level, using a 12 mm radius sphere centered in the peak of accuracy for classification of the valence of emotions in the study by Skerry and Saxe (2014), at MNI coordinates [−2, 50, 34]. Interactions between the ROIs' mean time-courses were modeled using the method described in sections 'Interaction Model' and 'Model Fitting' above. In Experiment 1, faces and situations were used as conditions, in Experiment 2, age task and valence task were used as conditions.

4.5. Preprocessing

Each participant's data were preprocessed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) and custom software written in Matlab (www.mathworks.com Natick, MA, USA). Each participant's data were registered to the first image of the first run, and all functional runs were co-registered with the participant's anatomical scan. All images (functional and anatomical) were normalized to a common brain space (Montreal Neurological Institute, EPI template).

5. Results

5.1. Experiment 1

In Experiment 1, the inter-regional influence parameters substantially improved the fit of the model, when added to the hierarchical model in the final stage. In the deconvolved timeseries, the influence parameters explained an additional 32.88% (SEM 2.02%) of the variance in left-out runs. A similar but weaker result was obtained without deconvolution, just predicting the original BOLD signal (additional variance 12.27%, SEM 0.9%, see Appendix A – Fig. 5). In light of the validation simulations, network structure (Fig. 4 A) was calculated modelling the deconvolved timeseries in the ROIs with a bilinear model and running a *t*-test on model parameters (see Materials and Methods for more details). Positive influences from OFA to FFA were observed ($t(25) = 3.24$, p=0.0033), along with negative influences in the opposite direction ($t(25) = -3.52$, p=0.0016). In addition to condition-independent influences, in Experiment 1 a significant increase in the influence from OFA to FFA was observed in the face condition ($t(25) = 2.28$, p=0.0314), but not in the situations condition ($t(25) = -0.10$, p=0.9197).

5.2. Experiment 2

In Experiment 2 the inter-regional influence parameters (Fig. 4B) explained substantial additional variance in the deconvolved timeseries (33.97%, SEM 1.35%) and also, to a lesser extent, in the original BOLD signal (17.4%, SEM 1.28%, see Appendix A – Fig. 5). These values are remarkably similar to the results of Experiment 1, suggesting that the importance of influence parameters is highly replicable, in this model, across participants and groups. To establish a baseline for the additional variance explained by influences between regions in deconvolved data, we ran an additional 1000 simulations. The connectivity matrices for the simulations were randomly generated for each subject and run using the same parameters as the DNM validation. We simulated 25 subjects and 4 ROIs, to match the numbers of participants and ROIs in our experiments. Across the 1000 simulations, the additional variance explained by the influences between regions was on average 4.84%, and it never equaled or exceeded 30%.

The similarity of the influence matrices between the two experiments was quantified using Pearson correlation: the correlation between the influence matrices in Experiment 1 and Experiment 2 was r=0.97. A direct comparison of the results in the two experiments using unpaired samples *t*-tests did not yield any significant differences. In Experiment 2 (Fig. 4), which requires task switching, significant influences were detected from MPFC to OFA and FFA ($t(24) = 2.15$, p=0.0416 and $t(24) = 2.28$, p=0.0316 respectively), and from pSTS to OFA ($t(24) = 2.82$, p=0.0095). A trend of influence from pSTS to FFA was also observed ($t(24) = 1.89$, p=0.0703). In addition to condition-independent influences, in Experiment 2 a significant decrease in the influence from MPFC to FFA was observed in the age task ($t(24) = -2.08$, p=0.0484), and a significant increase in the influence from MPFC to OFA ($t(24) = 2.47$, p=0.0209) and decrease from FFA to OFA were observed in the valence condition ($t(24) = -2.58$, p=0.0163).

6. Discussion

We have introduced and validated Dynamic Network Modelling (DNM), a method to investigate time-lagged influences between brain regions with fMRI. DNM is an intuitive and accurate framework for

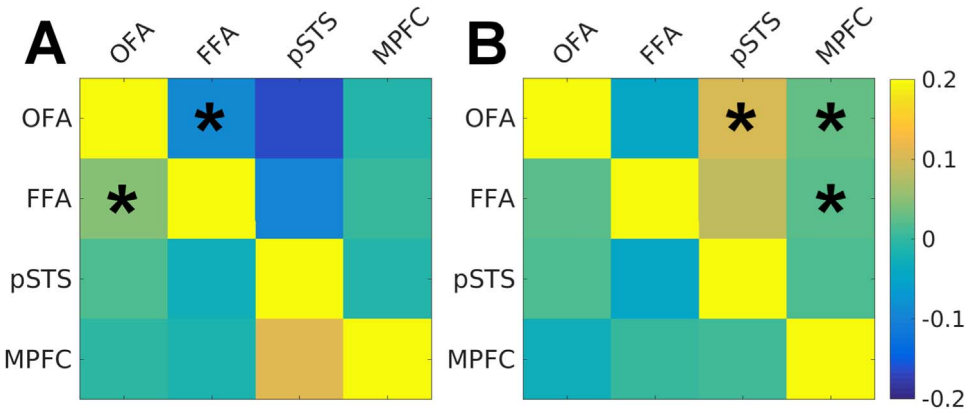

Fig. 4. (A) Color-coded matrix of parameter values for the influences between regions in Experiment 1; (B) color-coded matrix of parameter values for the influences between regions in Experiment 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

network discovery, with sensitivity to the sign as well as the value of influences between brain regions. The quality of fit of DNM is measured with variance explained in independent data, which can be used to compare different models and also as a warning if the variance explained by the models tested is too low.

We sought to measure directional, time-lagged influences between cortical regions, using functional MRI. An important challenge to using fMRI to study interations is the temporal resolution: the blood oxygenated dependent (BOLD) signal is inherently slow, and is typically measured at 0.5 Hz. Other neuroimaging techniques, including MEG and EEG offer much higher temporal resolution than fMRI, and therefore the potential to study interregional influences at the true frequency of neural computation (e.g. >100 Hz). On the other hand, fMRI has substantially better spatial resolution. Poor source reconstruction is a serious challenge for studying interregional influences, because misattributing distinct signals to a common source could create a false appearance of an influence where none exists. Also, over the past decade, the cognitive functions of many cortical regions, localized with fMRI, have been intensively investigated, in terms of the magnitude and pattern of responses to stimulus categories. Further theory building will require understanding how influences between these particular cortical regions support transformation of information (i.e. processing, Anzellotti et al., 2016a, 2016b).

Prior research investigating directional influences between brain regions in fMRI data has used two techniques: Granger Causality (GC, Roebroeck et al., 2005) and Dynamic Causal Modelling (DCM, Friston et al., 2003). While both techniques are powerful, and have been validated in some contexts (David et al., 2008), both also have key limitations, especially in the context of network discovery, as identified in the introduction. Here we developed a new approach, building on the strengths of both GC and DCM, but also designed to satisfy novel criteria. Like DCM, we used deconvolution to address potential differences in the haemodynamic responses across regions, and simultaneously estimated parameters for interregional influences, and condition-dependent influences. These approaches address some limitations of the most common implementation of GC. On the other hand, compared to DCM, GC is conceptually intuitive and computationally efficient. In our case, we used random effects statistical tests on VAR model parameters for efficient network discovery. Also, like GC, we aimed to provide a direct measure of variance explained, rather than the relative likelihood of a model (compared only to other models tested) provided by DCM.

Our approach (DNM) also goes beyond existing tools in multiple ways. First, we measure the variance explained by inter-regional influences, in independent left-out data. To measure the additional variance explained by inter-regional influences in a conservative manner, we used a step-wise regression, first greedily accounting for the stimulus-evoked responses and autocorrelations in the timeseries within region, before testing variance explained by inter-regional influences. In two experiments, we found that adding parameters for interregional influences to the model explained more than 30% of the remaining variance in the deconvolved timeseries of left-out runs. Evaluating variance explained in independent data is critical because it ensures that models are not overfitting the training data (variance explained in independent data is not biased in favor of more complex models). As a consequence, an explicit penalization of model complexity is not necessary. Using variance explained in independent data is appealing also because it provides an absolute (not relative) measure of the goodness of fit of these models.

Second, we conduct statistical tests directly on the model parameters, across subjects, to preserve information about the sign of inter-regional influences (whether they are excitatory or inhibitory). In the absence of a reliable influence between two regions, the parameter corresponding to that influence in the model will be highly variable across runs and participants, and will not reach significance. Validation results show that this approach is effective: DNM can accurately infer the presence and sign of inter-regional influences, with very few false positives (1).

In a set of simulations with varying numbers of participants, regions, and levels of noise, we have shown that DNM can accurately infer the presence as well as the sign of interactions between regions (Fig. 1). With increasing level of noise, DNM successfully maintained a high accuracy, erring on the conservative side by increasing the number of undetected influences. In a direct comparison with Granger Causality, we found that both methods were effective when influences between regions are consistent in sign across participants, but thanks to its sensitivity to the sign of influences, DNM, unlike GC, was able to discard the influences with highly variable sign across participants. This property of DCM can be an important asset when investigating heterogeneous populations, for example in a clinical setting.

We applied DNM to study influences between brain regions during emotion recognition in two experiments: an experiment on the recognition of emotions from videos of facial expressions and of situations, and an experiment on the recognition of the valence of the emotion and the age of a person from videos of facial expressions. In the two experiments, we found similar influences between brain regions involved in emotion recognition (r=0.97). In Experiment 1, we found a significant positive influence from OFA to FFA. A positive

influence from OFA to FFA also observed in Experiment 2. In the opposite direction, a significant negative influence from FFA to OFA was observed in Experiment 1. This negative FFA to OFA influence was also observed in Experiment 2. A negative influence from FFA to OFA can be due to OFA responses starting to decline earlier than FFA responses, so that an increase in FFA responses predicts an upcoming decline in OFA responses. This finding exemplifies the importance of distinguishing between positive and negative influences. In Experiment 1 the face condition, but not the situations condition, led to an even stronger influence from OFA to FFA, as is expected given the use of naturalistic face videos in the expression condition.

In Experiment 2, which required participants to switch tasks between recognizing the valence of a facial expression and the age of the person in the video, significant positive influences were observed from MPFC and pSTS, which are known to encode abstract representations of emotions (Skerry and Saxe, 2014) to ventral temporal face-selective regions (OFA, FFA). Given the greater ability of representations in MPFC and pSTS to generalize across different stimuli (Skerry and Saxe, 2014) than representations in OFA and FFA, we interpret these influences as top-down. A recent study (Kliemann et al., 2016) has reported task-dependent changes in the patterns of response in ventral temporal face-selective regions. Participants were asked to judge the age (in half of the trials) and the valence (in the remaining trials) of facial expressions, and found that patterns of response in the right FFA were more similar in the trials in which participants performed the same task than in trials in which they performed different tasks. By showing top-down influences from MPFC and pSTS to ventral temporal face-selective regions, the present results individuate a possible mechanism driving task-dependent changes in FFA during facial expression recognition.

In general, modulatory connections might be particularly amenable to detection by fMRI, since they could have relative slow temporal frequency, compared with bottom-up information processing. The temporal resolution of fMRI is low, therefore the physiological bases of the signal modeled in this study are probably not monosynaptic connections between neurons. Temporal variation in the BOLD signal is correlated with local field potentials, and more specifically with the power of the gamma band frequency oscillations and with slow cortical potentials (<4 Hz) (He et al., 2008). The power of gamma band frequency oscillations and slow cortical potentials account for different parts of the BOLD's variance (Scheeringa et al., 2011). Slow cortical potentials are mostly driven by synaptic activity at apical dendrites in superficial layers of cortex (Goldring, 1974; Mitzdorf, 1985; Birbaumer et al., 1990; He and Raichle, 2009), and activity in superficial layers depends on long-range connections (He and Raichle, 2009). High power in the gamma band usually co-occurs with gamma synchronization within a region (Fries, 2009). It reflects engagement of the region in a task (Pulvermüller et al., 1995) and may need to be triggered by 'modulatory network activation', for instance top-down control (Fries et al., 2001, 2008; Bichot et al., 2005; Taylor et al., 2005; Womelsdorf et al., 2006). Our results might depend on a combination of slow cortical potentials and gamma band power that reflect respectively low-frequency fluctuations in the activation state of a network and relatively faster sequences of responses due to engagement in a task.

Recent work has highlighted the importance of considering the multivariate structure of responses in the study of connectivity (Coutanche and Thompson-Schill, 2013; Geerligs and Henson, 2016; Anzellotti et al., 2016a, 2016b). Expanding DNM to incorporate a treatment of multivariate structure within ROI responses is a very promising avenue for future research.

In conclusion, we introduced an exploratory method (DNM) to model the directed and signed influences between brain regions while controlling for inter-regional differences in the haemodynamic responses; statistical tests on parameter values were used to perform efficient network discovery. We applied DNM to the study of the influences between brain regions encoding information about emotional valence during emotion recognition, showing how DNM has key assets as compared to other available methods and gaining new insights on the influences between brain regions processing emotional valence.

Acknowledgments

This study was supported by NIH Grant 1R01 MH096914-01A1 to Prof. Rebecca Saxe. Stefano Anzellotti was supported by a Postdoctoral Fellowship from the Simons Center for the Social Brain, Dorit Kliemann was supported by the Humboldt Foundation (Feodor Lynen Fellowship). We would like to thank the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research, MIT for providing scanning resources, and Rik Henson, Gabriele Anzellotti and Dimitris Pinotsis for comments on an earlier version of the manuscript.

Appendix A. Additional figures

(Fig. 5).

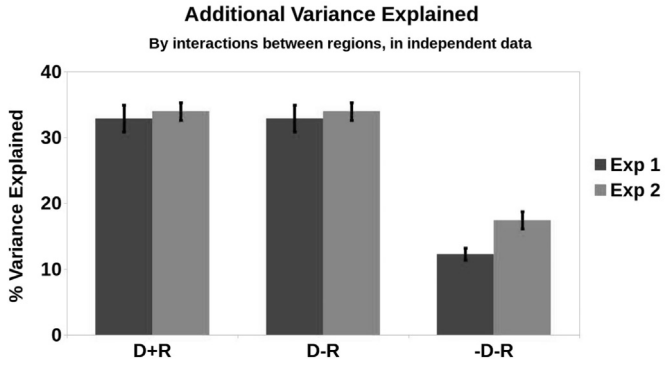

Fig. 5. Additional variance explained in independent data by the influences between regions. The model using both deconvolution and regularization is compared to a model using deconvolution but not regularization and to a model using neither deconvolution nor regularization.

Appendix B. Granger causality

Let $a(t)$, $b(t)$ with $t \in 1,\ldots,T$ be two timeseries, for example the timecourses of Blood-Oxygen Level Dependent (BOLD) signal in two brain regions A, B. We can consider the autoregressive model

$$b(t) = \beta b(t-1) + \epsilon_b \tag{3}$$

and the model

$$b(t) = \beta_1 a(t-1) + \beta_2 b(t-1) + \epsilon_{ab} \tag{4}$$

where ϵ_b and ϵ_{ab} are the residuals of the models. Here we focus for simplicity on a model between two regions that goes only 1 timestep in the past, but multiple regions and multiple timesteps in the past can also be considered. In Granger Causality, the influence from a brain region A to a brain region B is given by:

$$F_{A\rightarrow B} = \log\left(\frac{var(\epsilon_b)}{var(\epsilon_{ab})}\right). \tag{5}$$

If earlier responses in region A explain additional variance in the responses of region B, the ratio $var(\epsilon_b)/var(\epsilon_{ab})$ is greater than one; the logarithm remaps the range of possible values from $[1, +\infty)$ to $[0, +\infty)$. The suggested approach to control for differences in haemodynamic responses between regions consists in calculating the subtraction:

$$F^{12}_{A\rightarrow B} = F^{1}_{A\rightarrow B} - F^{2}_{A\rightarrow B} \tag{6}$$

where $F^{1}_{A\rightarrow B}$ is the influence from A to B during condition $C1$, and $F^{2}_{A\rightarrow B}$ is the influence from A to B during condition $C2$.

Appendix C. Dynamic causal modelling

Let $\mathbf{z} = [z_1(t),\ldots,z_n(t)]$ be a vector of neural responses at time t, where $z_i(t)$ is the response at time t in region i for each of n regions. In DCM, the change in neural responses $\dot{\mathbf{z}}(t)$ is modeled as

$$\dot{\mathbf{z}}(t) = A\mathbf{z}(t) + \sum_{j=1}^{m} u_j(t) B_j \mathbf{z}(t) + C\mathbf{u}(t) \tag{7}$$

where A is a matrix of fixed connectivity parameters, B_j is a matrix of connectivity parameters during condition j for each of the m conditions, and $u_j(t)$ is 1 if condition j is presented at time t and 0 otherwise. The matrix C contains parameters for the effect of the conditions $\mathbf{u}(t) = [u_1(t),\ldots,u_m(t)]$. A candidate model M is specified providing matrices of binary values $\widetilde{A}$, $\widetilde{B_j}$, $\widetilde{C}$ of the same size as A, B_j, and C, so that in M, parameter $A(h, k)$ is allowed to vary if and only if $\widetilde{A}(h, k) = 1$. The parameters of candidate models are estimated with the expectation-maximization (EM) algorithm, and the best model is chosen using Bayesian model comparison (Friston et al., 2003).

References

Anzellotti, S., Caramazza, A., 2015. From parts to identity: invariance and sensitivity of face representations to different face halves. Cereb. Cortex, bhu337.

Anzellotti, S., Caramazza, A., Saxe, R., 2016a. Multivariate pattern connectivity. bioRxiv, 046151.

Anzellotti, S., Fedorenko, E., Caramazza, A., Saxe, R., 2016b. Measuring and modeling transformations of information between brain regions with fMRI. bioRxiv.

Bichot, N.P., Rossi, A.F., Desimone, R., 2005. Parallel and serial neural mechanisms for visual search in macaque area v4. Science 308 (5721), 529–534.

Birbaumer, N., Elbert, T., Canavan, A.G., Rockstroh, B., 1990. Parallel and serial neural mechanisms for visual search in macaque area v4. Physiol. Rev. 70 (1), 1–41.

Bressler, S.L., Menon, V., 2010. Large-scale brain networks in cognition: emerging methods and principles. Trends Cogn. Sci. 14 (6), 277–290.

Buschman, T.J., Miller, E.K., 2007. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science 315 (5820), 1860–1862.

Chen, G., Hamilton, J.P., Thomason, M.E., Gotlib, I.H., Saad, Z., Cox, R., 2009. Multi-region granger causality tuned for fMRI data analysis. Annual Meeting of the International Society for Magnetic Resonance in Medicine.

Coutanche, M.N., Thompson-Schill, S.L., 2013. Informational connectivity: identifying synchronized discriminability of multi-voxel patterns across the brain. Front. Hum. Neurosci. 7, 15.

David, O., Guillemain, I., Saillet, S., Reyt, S., Deransart, C., Segebarth, C., Depaulis, A., 2008. Identifying Neural Drivers with functional MRI: An Electrophysiological Validation.

Davies-Thompson, J., Andrews, T.J., 2011. The localization and functional connectivity of face-selective regions in the human brain. J. Vis. 11 (11), (647–647).

Dodell-Feder, D., Koster-Hale, J., Bedny, M., Saxe, R., 2011. fMRI item analysis in a theory of mind task. NeuroImage 55 (2), 705–712.

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., Vuilleumier, P., 2011. Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. Cereb. Cortex, bhr113.

Ethofer, T., Bretscher, J., Wiethoff, S., Bisch, J., Schlipf, S., Wildgruber, D., Kreifelts, B., 2013. Functional responses and structural connections of cortical areas for processing faces and voices in the superior temporal sulcus. NeuroImage 76, 45–56.

Fedorenko, E., Thompson-Schill, S.L., 2014. Reworking the language network. Trends Cogn. Sci. 18 (3), 120–126.

Fries, P., 2009. Neuronal gamma-band synchronization as a fundamental process in cortical computation. Annu. Rev. Neurosci. 32, 209–224.

Fries, P., Reynolds, J.H., Rorie, A.E., Desimone, R., 2001. Modulation of oscillatory neuronal synchronization by selective visual attention. Science 291 (5508), 1560–1563.

Fries, P., Womelsdorf, T., Oostenveld, R., Desimone, R., 2008. The effects of visual stimulation and selective visual attention on rhythmic neuronal synchronization in macaque area v4. J. Neurosci. 28 (18), 4823–4835.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. NeuroImage 19 (4), 1273–1302.

Friston, K.J., Litvak, V., Oswal, A., Razi, A., Stephan, K.E., van Wijk, B.C., Ziegler, G., Zeidman, P., 2016. Bayesian model reduction and empirical bayes for group (dcm) studies. NeuroImage 128, 413–431.

Friston, K., Penny, W., 2011. Post hoc Bayesian model selection. Neuroimage 56 (4), 2089–2099.

Furl, N., Hadj-Bouziane, F., Liu, N., Averbeck, B.B., Ungerleider, L.G., 2012. Dynamic and static facial expressions decoded from motion-sensitive areas in the macaque monkey. J. Neurosci. 32 (45), 15952–15962.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., Benedetti, F., Abbamonte, M., Gasparotti, R., Barale, F., et al., 2009. Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. J. Psychiatry Neurosci. 34 (6), 418.

Gallagher, H.L., Frith, C.D., 2003. Functional imaging of 'theory of mind'. Trends Cogn. Sci. 7 (2), 77–83.

Gauthier, I., Tarr, M.J., Moylan, J., Skudlarski, P., Gore, J.C., Anderson, A.W., 2000. The fusiform "face area" is part of a network that processes faces at the individual level. J. Cogn. Neurosci. 12 (3), 495–504.

Geerligs, L., Henson, R.N., et al., 2016. Functional connectivity and structural covariance between regions of interest can be measured more accurately using multivariate distance correlation. NeuroImage 135, 16–31.

Goldring, S., 1974. Dc shifts released by direct and afferent stimulation. Handb. Electroencephalogr. Clin. Neurophysiol. 10 (Part A), 12–24.

Gschwind, M., Pourtois, G., Schwartz, S., Van De Ville, D., Vuilleumier, P., 2011. White-matter connectivity between face-responsive regions in the human brain. Cereb. Cortex, bhr226.

Hamilton, J.P., Chen, G., Thomason, M.E., Schwartz, M.E., Gotlib, I.H., 2011. Investigating neural primacy in major depressive disorder: multivariate granger causality analysis of resting-state fMRI time-series data. Mol. Psychiatry 16 (7),

763–772.
Hariri, A.R., Bookheimer, S.Y., Mazziotta, J.C., 2000. Modulating emotional responses: effects of a neocortical network on the limbic system. NeuroReport 11 (1), 43–48.
He, B.J., Raichle, M.E., 2009. The fMRI signal, slow cortical potential and consciousness. Trends Cogn. Sci. 13 (7), 302–309.
He, B.J., Snyder, A.Z., Zempel, J.M., Smyth, M.D., Raichle, M.E., 2008. Electrophysiological correlates of the brain's intrinsic large-scale functional architecture. Proc. Natl. Acad. Sci. USA 105 (41), 16039–16044.
Ishai, A., 2008. Let's face it: it's a cortical network. NeuroImage 40 (2), 415–419.
Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17 (11), 4302–4311.
Kliemann, D., Jacoby, N., Anzellotti, S., Saxe, R.R., 2016. Decoding task and stimulus representations in face-responsive cortex. Cogn. Neuropsychol., 1–16.
Mitzdorf, U., 1985. Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. Am. Physiol. Soc..
Narumoto, J., Okada, T., Sadato, N., Fukui, K., Yonekura, Y., 2001. Attention to emotion modulates fMRI activity in human right superior temporal sulcus. Cogn. Brain Res. 12 (2), 225–231.
Peelen, M.V., Atkinson, A.P., Vuilleumier, P., 2010. Supramodal representations of perceived emotions in the human brain. J. Neurosci. 30 (30), 10127–10134.
Pinotsis, D.A., Perry, G., Litvak, V., Singh, K.D., Friston, K.J., 2016. Intersubject variability and induced gamma in the visual cortex: DCM with empirical bayes and neural fields. Hum. Brain Mapp.. http://dx.doi.org/10.1002/hbm.23331, (n/a-n/a. ISSN 1097-0193).
Pitcher, D., Dilks, D.D., Saxe, R.R., Triantafyllou, C., Kanwisher, N., 2011. Differential selectivity for dynamic versus static information in face-selective cortical regions. NeuroImage 56 (4), 2356–2363.
Pulvermüller, F., Lutzenberger, W., Preil, H., Birbaumer, N., 1995. Spectral responses in the gamma-band: physiological signs of higher cognitive processes? NeuroReport 6 (15), 2059–2064.
Pyles, J.A., Verstynen, T.D., Schneider, W., Tarr, M.J., 2013. Explicating the face perception network with white matter connectivity. PLoS One 8 (4), e61611.
Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using granger causality and fMRI. NeuroImage 25 (1), 230–242.
Scheeringa, R., Fries, P., Petersson, K.-M., Oostenveld, R., Grothe, I., Norris, D.G., Hagoort, P., Bastiaansen, M.C., 2011. Neuronal dynamics underlying high-and low-frequency eeg oscillations contribute independently to the human bold signal. Neuron 69 (3), 572–583.
Sergent, J., Ohta, S., Macdonald, B., 1992. Functional neuroanatomy of face and object processing. Brain 115 (1), 15–36.
Skerry, A.E., Saxe, R., 2014. A common neural code for perceived and inferred emotion. J. Neurosci. 34 (48), 15997–16008.
Spunt, R.P., Lieberman, M.D., 2012. An integrative model of the neural systems supporting the comprehension of observed emotional behavior. NeuroImage 59 (3), 3050–3059.
Stephan, K.E., Penny, W.D., Moran, R.J., den Ouden, H.E., Daunizeau, J., Friston, K.J., 2010. Ten simple rules for dynamic causal modeling. NeuroImage 49 (4), 3099–3109.
Taylor, K., Mandon, S., Freiwald, W., Kreiter, A., 2005. Coherent oscillatory activity in monkey area v4 predicts successful allocation of attention. Cereb. Cortex 15 (9), 1424–1437.
Turk-Browne, N.B., Norman-Haignere, S.V., McCarthy, G., 2010. Face-specific resting functional connectivity between the fusiform gyrus and posterior superior temporal sulcus. Front. Hum. Neurosci. 4.
Winston, J.S., Gottfried, J.A., Kilner, J.M., Dolan, R.J., 2005. Integrated neural representations of odor intensity and affective valence in human amygdala. J. Neurosci. 25 (39), 8903–8907.
Womelsdorf, T., Fries, P., Mitra, P.P., Desimone, R., 2006. Gamma-band synchronization in visual cortex predicts speed of change detection. Nature 439 (7077), 733–736.
Woolrich, M.W., Ripley, B.D., Brady, M., Smith, S.M., 2001. Temporal autocorrelation in univariate linear modeling of fMRI data. NeuroImage 14 (6), 1370–1386.
Zaki, J., Ochsner, K., 2009. The need for a cognitive neuroscience of naturalistic social cognition. Ann. N. Y. Acad. Sci. 1167 (1), 16–30.